NLUAstro 1, 39-58 (2019) Notices of Lancaster University Astrophysics PHYS369: Astrophysics Group project

High Redshift LAEs and their Cosmic Evolution: Morphologies, SFR and AGN Activity from z ∼ 2 to 6⋆

Cassandra Barlow-Hall[1]†, Joseph Bramwell[1], Daniel Hodder[1], Michael Merrett[1], Adam Russ[1], Oliver Wareing[1] & David Sobral[1]‡
[1] *Department of Physics, Lancaster University, Lancaster, LA1 4YB, UK*

Accepted 29 May 2019. Received 29 May 2019; in original form 25 March 2019

ABSTRACT
We studied a large sample of ∼ 4000 high redshift Lyman-alpha Emitters (LAEs) in order to determine their properties and infer how they might have evolved into the local Universe. This was done through the exploration of the SC4K survey (Sobral et al. 2018a) and making use of the *Hubble Space Telescope* (*HST*), *Chandra X-ray Observatory* (*Chandra*) and the Very Large Array (VLA). We find that SC4K LAEs are mostly (69 ± 4%) compact disky galaxies (average Sérsic index, $n = 1.9 \pm 2.2$) The average star formation rate $SFR_{Ly\alpha}$ of LAEs is $\approx 17\,M_\odot\,yr^{-1}$. We find that SFR increases with increasing stellar mass. We also observed a characteristic 'peak' in SFR at $M \sim 10^{9.3}\,M_\odot$, at redshift $z \sim 2.5$, and progressing to higher stellar masses at higher redshifts. We find a total of 303 X-ray or radio detected active galactic nuclei (AGN) within the SC4K catalogue. These AGN have a range of black hole accretion rates (BHARs) from $\sim 0.03\,M_\odot\,yr^{-1}$ to $\sim 3.3\,M_\odot\,yr^{-1}$. The AGN fraction increases with increasing Lyα luminosity and decreases with increasing redshift, peaking at $z \sim 3$. LAEs found at $z \sim 2-6$ with a stellar mass $M \sim 10^{10}\,M_\odot$ and a SFR $\sim 5.4\,M_\odot yr^{-1}$ are consistent with being progenitors of Milky Way-like galaxies progenitor. Additionally, we found that the majority of the SC4K LAEs consists of cluster-like progenitors that will go on to form the brightest cluster galaxies (BCGs) in the local Universe.

Key words: Lyα Emitters, Galaxies, Star Formation, Stellar Mass, Progenitors.

1 INTRODUCTION

Galaxies formed in the early Universe, a few billion years ago, as matter cooled and collapsed into regions of higher density, described by Λ cold dark matter (ΛCDM) cosmology. Baryonic matter coalesced into stars which then became gravitationally bound into galaxies which where surrounded by dark matter halos that formed the cosmic web (see Somerville & Dave (2015) and references therein). These first galaxies continued to evolve, with stars dying and enriching the surrounding interstellar space with gas and dust. New stars then form as the gas and dust cools and collapses into denser regions once again.

With the advancements in high redshift observations, understanding the formation and evolution of early galaxies over cosmic time is the focus of much research (see e.g. Stark 2016; Gruppioni et al. 2013; Lehmer et al. 2013a; Marchesini et al. 2009; Muzzin et al. 2013; van Dokkum et al. 2013).

As the very early Universe contained no metals (atoms and ions above a proton number of 2), the first stars that formed where comprised of only hydrogen and helium. These first stars, called Population III stars, were very massive and short lived, forming metals and enriching galaxies when they died (e.g Sobral et al. 2015). The stars that followed, Population I and Population II, did contain metals and thus where smaller with longer lifetimes, as fusion within these stars could proceed through the more efficient CNO cycle, allowing the stellar mass of galaxies to increase.

Galaxies continue to form stars as they evolve through cosmic time. Many studies have shown that the star formation rate density (SFRD) of the Universe evolves with cosmic time, peaking at a redshift of $z \sim 2-3$ (e.g. Lilly et al. 1996; Sobral et al. 2012b; Madau & Dickinson 2014) and decreasing towards higher redshifts, and thus earlier cosmic times (see Khostovan et al. 2015).

The hydrogen spectral emission line Lyman-alpha (Lyα), with a rest-frame wavelength of 1215.67 Å, is found

⋆ Based on SC4K (Sobral et al. 2018a) and on observations obtained with the *HST*, Subaru, VLA, *Chandra* and INT.
† E-mail: c.barlow-hall@lancaster.ac.uk
‡ PHYS369 supervisor.

to be strongly emitted by star forming galaxies (SFGs) and active galactic nuclei (AGN) (e.g. Charlot & Fall 1993) and is the strongest optical and UV emission line, with the high equivalent width (EW_0) making spectroscopic follow-up easy for galaxies that emit Lyα photons (e.g. Hashimoto et al. 2017). However Lyα photons are readily scattered by the gas and dust in the interstellar medium (ISM) of the galaxies that produced them and in the circumgalactic medium (CGM) surrounding galaxies (see Wisotzki et al. 2016). This leads to changes in the strength of the Lyα line of a galaxy's emission spectrum, depending upon the relative distribution of gas and dust along the line of the emission path between the source and the observer (e.g. Laursen et al. 2013; Neufeld 1991). As such not all star forming galaxies can be identified with this spectral line.

Lyα emitting galaxies (LAEs) have been found to be young, bluer galaxies with a low dust content (e.g. Oteo et al. 2015; Sobral et al. 2015; Bacon et al. 2015), as Lyα typically probes lower stellar mass galaxies (e.g. Hagen et al. 2016). Other studies, however, suggest Lyα probes more varied types of galaxies, with some old, redder, dusty LAEs being found (e.g. Hathi & Le Fèvre 2016). Most early SFGs are found to be LAEs with progenitors of many types of present day galaxies (e.g. Pirzkal et al. 2007; van Dokkum et al. 2013), making them excellent sources from which the evolutionary history of present day galaxies can be inferred.

As Lyα traces the formation of stars in a galaxy, the Lyα luminosity of a galaxy is regularly used to calculate the SFR of LAEs (see e.g. Sobral et al. 2018a, and references there in). After the epoch of re-ionisation, at $z \sim 6$, more light was able to escape from galaxies, producing an increase in the number of LAEs which are observed (e.g Hayes et al. 2011). The star formation rate (SFR) of galaxies after this redshift has been seen to remain approximately constant, with a characteristic Lyα luminosity of $10^{42.93}$ erg s^{-1} up to $z \sim 2$ (e.g. Sobral et al. 2018a). Other studies (e.g. Santos et al. 2016) have also shown that the function of the Lyα emission remains approximately constant over a redshift range of $z \sim 3-6$. The SFR of a galaxy can be calculated from the measured luminosity of the galaxy at a particular wavelengths which are emitted by short lived stars or by the gas around star forming regions, with UV, far infrared (FIR), radio and X-ray wavelengths being among those used.

Despite the majority of the population of LAEs being SFGs, the brightest of these galaxies are typically found to be AGN (e.g. Sobral et al. 2018b). Thus, the Lyα emission observed for these galaxies may be produced by the supermassive black holes at the centre of these AGN and therefore Lyα traces the accretion rate of AGN (e.g. Calhau 2019). AGN also appear as bright X-ray sources, due to the X-ray emissions of the black hole accretion disc, due to the frictional heating of matter within the accretion disc, with most X-ray detections occurring in a redshift range of $z \sim 2.2-3.5$ (e.g. Haardt & Maraschi 1991). This allows the black hole accretion rate (BHAR) to be calculated from the X-ray luminosity, which provides a measure of the activity of the AGN. Using the X-ray emission of AGN, the BHAR has been found to evolve over time, peaking at a similar redshift to the SFRD at a redshift of $z \sim 2-3$ (e.g Madau & Dickinson 2014; Lehmer et al. 2013b).

AGN have also been found to emit strongly in radio wavelengths. Unlike the X-ray and Lyα emission, the radio emission of AGN traces the high energy particle jets and lobes that form from particles ejected by the supermassive black hole into the intergalactic medium (IGM) (e.g. Netzer 1989). Studies have found no relation between the Lyα and radio emission of AGN, which is likely due to the different mechanisms of AGN activity which the two emissions trace (e.g. Calhau 2019).

The morphology and size of galaxies has been found to evolve over cosmic time, becoming larger and more structured towards lower redshifts. LAEs are typically compact galaxies at higher redshifts (see Paulino-Afonso et al. 2018, and references there in), with SFGs also being found to have similar morphologies in the early Universe. Studies have also shown that the most massive galaxies are far more compact at around cosmic noon, $z \sim 2$, (e.g Cassata et al. 2010; van Dokkum et al. 2013) which leads us to expect there to be some dramatic increase in the overall size of these galaxies with little additional formation of stellar mass between $z \sim 2$ and $z \sim 0$ (see e.g. Vika et al. 2013, and references therein). This increase in size could be due to the rapid production of stellar mass prior to this increase in size, causing these galaxies to rapidly expand due to the sudden increase in radiation pressure between the stars.

In Paulino-Afonso et al. (2018), with only a small sample size, little evolution was found in the size of LAEs since a redshift of $z \sim 6$, however some evidence for size evolution at a redshift of $z \sim 2.5$. In the same study, the Sérsic indices, a measure of the distribution of light emission from which a measure of the radius of a galaxy can be found, (see e.g. Sérsic 1963; Caon et al. 1993), of LAEs were found to evolve with cosmic time, irrespective of the Lyα luminosity of the galaxy itself.

The evolution of the SFR, morphology and AGN activity of LAEs with cosmic time, and their growth into the galaxies in our local neighbourhood is the focus of this paper. We investigate properties of the SC4K sample, presented in Sobral et al. (2018a), which is comprised of 3908 LAEs across a redshift range of $z \sim 2-6$. This is done using catalogues of spectral data for these Lyα galaxies and images from the COSMOS Survey (see Elvis et al. 2009), from which the morphologies, SFR and AGN activity for this catalogue of LAEs are measured and the growth of the Lyα galaxies and evolutionary trends in the data investigated.

Our paper is organised as follows. In Section 2 we present the SC4K catalogue used in this paper. The methods employed are presented in Section 3, with the methods for investigating the morphologies, AGN activity and SFR detailed in Section 3 and the results presented in Section 4. The methodology and results on progenitor LAEs are presented in Sections 5.1 and 5.2. The results then discussed in Section 6 and finally the conclusions are presented in Section 7. Throughout this paper we use AB magnitudes (Oke & Gunn 1983), Salpeter IMF (Salpeter (1955)) and ΛCDM cosmology, with $H_0 = 70.0$ kms^{-1}Mpc^{-1}, $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$.

2 DATA AND SAMPLE

2.1 The SC4K Sample of Lyman-alpha Emitters

In this work we explore the SC4K sample which contains $\sim$ 4000 LAEs with redshifts ranging from $z \sim 2-6$ split

between 16 redshift slices, located within the COSMOS field, as in Sobral et al. (2018a). These LAEs were selected by identifying objects with an emission line equivalent width (EW) $> 50\times(1+z)$ Å and an emission line excess significance (Σ) > 3. EW was calculated in Sobral et al. (2018a) as:

$$\mathrm{EW} = \Delta\lambda_{\mathrm{MB}} \frac{f_{\mathrm{MB}} - f_{\mathrm{BB}}}{f_{\mathrm{BB}} - f_{\mathrm{MB}}(\Delta\lambda_{\mathrm{MB}}/\Delta\lambda_{\mathrm{BB}})} \quad (1)$$

where f_{MB} and f_{BB} are the flux densities of the filters and $\Delta\lambda_{\mathrm{MB}}$ and $\Delta\lambda_{\mathrm{BB}}$ are the full width at half maximum (FWHM) of the medium band (MB) and broad band (BB) filters (see Table 3).

This catalogue incorporates data from 12 MB and 2 NB filters from the Suprime-Cam instrument on the Subaru telescope, complemented by 2 narrow band (NB) filters from the 2.5 m Isaac Newton Telescope's (INT) Wide Field Camera. Between these 16 filters the data encompasses a range of wavelengths from $3920-8270$ Å(Sobral et al. 2018a). The 5σ depth in 2.1 arcsecond apertures of the MB filters ranges from $25.4-26.1$, whilst the 5σ depth in 3 arcsecond apertures of the NB filters ranges from $23.7-24.5$. The precise methodology used to obtain this catalogue is detailed in Sobral et al. (2018a). Other studies in the literature have used this sample e.g. Sobral et al. (2018a), Shibuya et al. (2019).

2.2 Obtaining Images for Visual Inspection

The images used throughout this study were obtained using the COSMOS Cutouts tool[1], which provides fields with $1-180$ arcsecond diameters within the COSMOS Survey field as seen by a variety of telescopes and filters. These images are provided by the COSMOS archive, which serves the Cosmic Evolution Survey project – a targeted survey covering a 2 deg^2 area over the equator with a wide variety of ground and space-based telescopes. Between these telescopes this field is imaged in a variety of wavelengths, ranging from X-ray to radio. In this paper we used images from the *HST* ($\lambda = 8140$ Å) (Koekemoer et al. 2007), *Chandra* ($\lambda = 2-25$ Å) (Civano et al. 2016) and VLA (λ = 10 cm, 20 cm) (Smolčić et al. 2017).

3 METHODOLOGY

3.1 Morphologies

3.1.1 Visual Classifications

A set of $2''$ by $2''$ images of the 1000 LAEs with the lowest broad band magnitudes, M_{BB}, in the SC4K data set was produced using the the *HST*-ACS Tiles setting on the COSMOS Cutouts tool, as mentioned in Section 2.2. This reduced set was used because many of the 3908 SC4K galaxies cannot be visually classified using *HST*-ACS Tiles images as they are not visible in said images. Thus we only classified galaxies with $M_{\mathrm{BB}} < 25.1974$ (this number was arbitrarily chosen to give a set of 1000 galaxies). Broad band magnitude here refers to use of the F814W filter which has an effective wavelength, $\lambda_{eff} = 8140$ Å, and a rest frame that is dependant upon the relationship, $\frac{8140}{1+z}$, where z is the redshift of the source, such that the rest frame at our highest redshift ($z \sim 5.8$) is ~ 1197 Å and at the lowest ($z \sim 2.2$) is ~ 2528 Å. Each of the galaxies in this reduced data set (known as 1000 Brightest, see Table 1) was then independently classified by three different team members in SAOImage DS9, using the same viewing settings. The classification scheme used split galaxies into four groups of decreasing compactness (inspired by a similar method used in Paulino-Afonso et al. 2018), with 0 as point-like, 1 as compact, 2 as disky and 3 as an irregular/merger galaxy as shown in Figure 1. The additional classification of -1 was added for galaxies it was not possible to classify on this scale, either because there was no image of the galaxy or the galaxy could not be seen in the image.

A mean was calculated from the three independent classifications as well as a standard deviation. The independent classifications were then plotted against this mean to ensure that, while independent, they were consistent with each other. This showed that each set of classifications had a correlation of greater than 0.9 (with 1 being perfect correlation) with the mean and thus these results could be used for further analysis. A reduction condition was applied that all means in the range of -1 to -0.5 were removed as they could not be classified and then the mean visual classification and standard deviation per redshift slice was calculated. For a complete list of data reduction conditions and distribution of sources with redshift see Table 1.

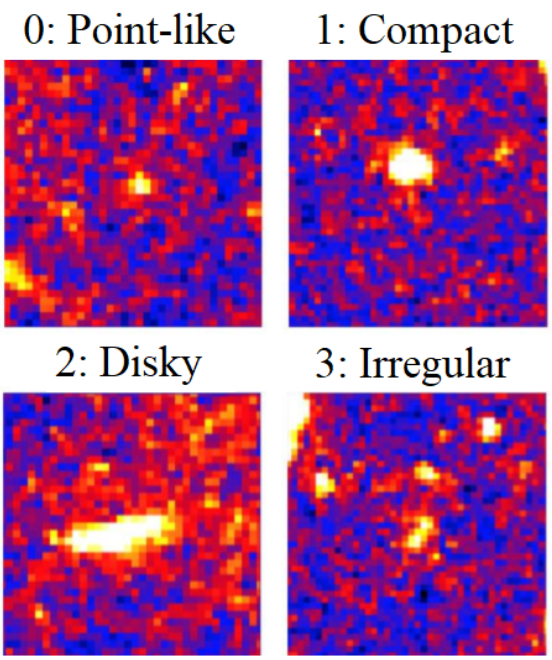

Figure 1. Example galaxies and their assigned visual classifications. Top left: 0 corresponds to point-like, a small point source. Top right: 1 corresponds to compact, an extended circular source. Bottom left: 2 corresponds to disky, an extended oval or disk-shaped source. Bottom right: 3 corresponds to irregular, a clumped obscurely-shaped source with multiple components. There is no example of the -1 classification as these were unclassifiable and as such do not have a set morphology.

[1] https://irsa.ipac.caltech.edu/data/COSMOS/index_cutouts.html

3.1.2 GALFIT Analysis of SC4K LAEs

We use results from Paulino-Afonso et al. (2018), in which the Sérsic profile was fitted to the observed light profile of all of the 3908 galaxies in the SC4K data set using GALFIT (Peng et al. 2002, 2010). This produced values for the fitted half light radius (r_e), Sérsic index (n), observed 20%, 50% and 80% light radii (r_{20}, r_{50} and r_{80}, the radii at which 20%, 50% and 80% of the observed light falls within r, respectively), and the compactness (C) of 3081 galaxies. While GALFIT converged for 3081 of the 3908 SC4K galaxies, a further 1478 r_e and n values were removed under the conditions that $\frac{1}{3} < \frac{r_e}{r_{50}} < 3$, which ensures r_e is kept relatively close to r_{50}, and that $n < 10$, as where this limit is exceeded it is an indication that the fitting process has failed (the resulting distribution of the data set can be seen in Table 1). The remaining results were then used both in this form and, as in Section 3.1.1, in a mean value per redshift slice (alongside a corresponding standard deviation).

3.1.3 Stellar Masses

In Santos et al. (in prep.) different magnitudes are transformed into flux densities, SED fitting is then used to fit this data with curves describing stellar population from which the stellar mass needed to produce said populations were then calculated. This paper uses a reduced set of this data by removing all AGN (by removing all X-ray and radio sources, see Section 3.2), which skew stellar mass measurements due to their extra light output, and galaxies with stellar masses M such that $M_{stellar} < 8$, where $M_{stellar} = \log_{10}\left(\frac{M}{M_\odot}\right)$. This is because these values are numerical artefacts due to our data not being deep enough to detect such galaxies as they are not luminous enough. The mean stellar mass per redshift slice and corresponding standard deviation was then calculated from the resulting data set (see Table 1) as in Section 3.1.1.

3.2 X-ray and Radio: AGN

3.2.1 Determining AGN

An image of the entire COSMOS field was obtained as viewed by the *Chandra X-ray Telescope* in three different X-ray energy bands: $0.5-2$ keV (soft), $2-7$ keV (hard) and $0.5-7$ keV (full). As in Sobral et al. (2018a), we follow (Calhau 2019) and create cutouts of each object to then calculate their X-ray fluxes, X-ray luminosities and black hole accretion rates (BHAR). The same method was then used on similar whole field images from the VLA radio observatory, with bands of 1.4 GHz and 3.0 GHz. The two new catalogues were then matched together, and the 303 objects which had been detected in any of the wavebands were marked as AGN. Note that there is some overlap - some objects were detected in multiple wavebands. An object was considered detected in a band if its signal was greater than 3 times the noise value, and the number of sources that fit this criteria in each band are given in Table 2.

Each of these was visually inspected to determine which had visible jets, lobes or particularly bright centres in the X-ray and radio wavelengths. This was done by viewing a 15 arcsecond cutout image of each object in both radio wavelengths (10 cm and 20 cm) and the soft X-ray energy band ($0.5-2$ keV), then combining these three images into one, with each wavelength being assigned to be either red, green or blue for added clarity. These colour images were then inspected to identify key features. This process also revealed that two of the objects identified as AGN were the same object viewed in the SC4K sample through different filters, as explained in Sobral et al. (2018a).

3.2.2 Calculating Black Hole Accretion Rate

To calculate the BHAR of the AGN, we follow Calhau (2019), beginning with calculating the X-ray flux from the count rate, as in Equation 2.

$$F_{X\text{-ray}} = (\text{counts/s}) \times CF \times 10^{-11} (\text{erg s}^{-1}\text{cm}^{-2}) \quad (2)$$

CF is a conversion factor with values of 0.687, 3.05 and 1.64 for the soft, hard and full X-ray bands respectively. These values were obtained by taking the average of the values given in Elvis et al. (2009) and the more recent paper Civano et al. (2016). From these fluxes we then calculate the luminosity with Equation 3,

$$L_{X\text{-ray}} = 4\pi F_{X\text{-ray}} d_L^2 (\text{erg s}^{-1}) \quad (3)$$

where d_L is the luminosity distance in cm, converted from the values given in Table 3.

This then required correction by multiplying it by a K-correction factor and converting each band's luminosity into the $0.5-10$ keV luminosity, as in Equation 4:

$$L_{0.5-10\text{keV}} = \frac{L_{X\text{-ray}}(10^{(2-\Gamma)} - 0.5^{(2-\Gamma)})}{(E_{max}(1+z))^{(2-\Gamma)} - (E_{min}(1+z))^{(2-\Gamma)}} \quad (4)$$

where z is the redshift, Γ is the photon index assumed to be 1.4 as in Calhau et al. (in prep.) and E_{max} and E_{min} are the maximum and minimum energies of each band.

Before obtaining the BHAR, we must multiply our K-corrected luminosities by a bolometric correction value. This value is extremely variable, but we take it to be 22.4 as in Lehmer et al. (2013a), as this is the median value for AGN with $L_{X\text{-ray}} = 10^{41} - 10^{46}$ erg s^{-1}. The BHAR of our sources can then be calculated using Equation 5:

$$\dot{M}_{BH} = \frac{(1-\epsilon)L_{bol}^{AGN}}{c^2 \epsilon} \quad (5)$$

where $\dot{M}_{BH}$ is the BHAR in units of $M_\odot yr^{-1}$, c is the speed of light and ϵ is the accretion efficiency, taken to be 0.1 as in Marconi et al. (2004).

3.3 Calculating Star Formation Rates

There are a number of techniques and calibrations one can use to estimate the star formation rate (SFR) in distant galaxies. HII regions are created around O and B type stars when ultraviolet (UV) photons ionise the surrounding hydrogen. When the ionised hydrogen ions recombine with electrons, they emit photons of characteristic frequency such as Lyα and H-alpha (Hα). In this way we can measure the number of photons and calculate the SFR in distant galaxies. Ideally we would use the Hα line to measure instantaneous SFR. However, the Hα rest frame is at 6563 Å, which is no longer visible in the optical range at redshifts greater than

Data Set	Reduction Conditions	Total Number of Sources	Data Set Used for
Full SC4K	None	3908	
1000 Brightest	$M_{\rm BB} < 25.1974$	1000	Cutouts
VC Mean Trusted	$M_{\rm BB} < 25.1974$, VC Mean > -0.5	788	VC Mean
GALFIT Output	GALFIT Converges	3081	r_{20}, r_{50}, r_{80}, C
Sérsic Trusted	GALFIT Converges, $\frac{1}{3} < \frac{r_e}{r_{50}} < 3$, $n < 10$	1603	r_e, n
Mass Trusted	$\mathrm{M_{stellar}} > 8$	3425	$\mathrm{M_{stellar}}$

Redshift z	Full SC4K Sources (%)	VC Mean Trusted Sources (%)	GALFIT Output Sources (%)	Sérsic Trusted Sources (%)	Mass Trusted Sources (%)
2.22	4.1 ± 0.3	5.8 ± 0.9	4.3 ± 0.4	3.4 ± 0.5	1.8 ± 0.2
2.51	19.0 ± 0.8	18.8 ± 1.7	19.4 ± 0.9	20.1 ± 1.2	19.6 ± 0.8
2.82	8.0 ± 0.5	11.5 ± 1.3	7.7 ± 0.5	7.6 ± 0.7	8.3 ± 0.5
2.98	18.2 ± 0.7	11.3 ± 1.3	18.5 ± 0.8	18.3 ± 1.2	18.9 ± 0.8
3.12	1.15 ± 0.17	2.2 ± 0.5	1.4 ± 0.2	1.1 ± 0.3	0.76 ± 0.15
3.15	12.4 ± 0.6	18.4 ± 1.7	12.3 ± 0.7	13.3 ± 1.0	13.2 ± 0.7
3.33	16.4 ± 0.7	15.5 ± 1.5	16.6 ± 0.8	16.8 ± 1.1	17.3 ± 0.8
3.72	2.5 ± 0.3	4.6 ± 0.8	2.5 ± 0.3	2.7 ± 0.4	2.7 ± 0.3
4.13	3.6 ± 0.3	1.8 ± 0.5	3.7 ± 0.4	3.8 ± 0.5	3.9 ± 0.3
4.58	2.0 ± 0.2	1.5 ± 0.4	2.2 ± 0.3	2.1 ± 0.4	1.9 ± 0.2
4.83	2.1 ± 0.2	2.9 ± 0.6	2.0 ± 0.3	1.9 ± 0.3	2.3 ± 0.3
4.85	2.0 ± 0.2	1.3 ± 0.4	1.9 ± 0.3	1.9 ± 0.4	2.1 ± 0.2
5.07	2.0 ± 0.2	2.8 ± 0.6	2.0 ± 0.3	2.0 ± 0.4	2.0 ± 0.2
5.31	0.84 ± 0.15	1.0 ± 0.4	0.78 ± 0.16	0.7 ± 0.2	0.88 ± 0.16
5.71	4.9 ± 0.4	–	4.8 ± 0.4	3.6 ± 0.5	3.8 ± 0.3
5.80	0.90 ± 0.15	0.5 ± 0.3	0.75 ± 0.16	0.50 ± 0.18	0.67 ± 0.14

Table 1. The reduction conditions used in Section 3.1 on the SC4K catalogue and the number of sources in the data sets produced, including the percentage of each data set's total number of sources in each redshift slice $\left(\text{Percentage} = \frac{\text{Number of Trusted Sources in Slice}}{\text{Total Number of Sources for this Data Set}}\right)$. Where $M_{\rm BB}$ is the broad band magnitude of the galaxy (note, this will range from u and B to i and z bands) and VC Mean is its mean visual classification. The – in the VC Mean Trusted column is due to a lack of any sources for this redshift slice.

Waveband	Number of Detected Sources	Detected Sources (%)
1.4GHz	54	1.38
3.0GHz	84	2.15
Soft	235	6.01
Hard	206	5.27
Full	223	5.71

Table 2. The number of sources with a signal value greater than 3 times that of the noise value in each waveband. 1.4GHz and 3.0GHz are the radio wavebands, and Soft, Hard and Full are the X-ray wavebands with energies of $0.5-2$ keV, $2-7$ keV and $0.5-7$ keV respectively as described in Section 3.2.1.

$z \sim 2.5$, and hence we cannot observe it using ground based telescopes. Future work (when the JWST launches) will be able to calculate SFRs at high redshift using the Hα line, which can then be compared to that of Lyα.

In this paper, we use Lyα luminosities, rest-frame UV magnitudes, along with radio and far infrared (FIR) stacking to calculate SFRs. Each of these processes traces a different aspect of star formation, and thus gives different implications for SFRs.

The first step in making use of the data was to reduce the relevant catalogues to only those sources not affected by AGN. This is because the accretion disks of AGN emit photons across a range of frequencies, some frequencies to a greater extent than others (e.g. radio and Lyα) – see Section 4.2 – and would therefore produce seemingly high SFR values and contaminate the results if included. We eliminated all possible AGN by excluding all sources that were detected to emit significantly in either radio and/or X-ray (as detailed in Section 3.2), which are clear indicators for AGN activity. Despite the FIR data not being so affected by AGN emissions – as FIR traces dust-obscured SFRs (Calhau et al. in prep.) – excluding them meant that the subsequently obtained SFRs from FIR stacking could be compared alongside those from radio.

The methodologies behind Lyα luminosity and UV magnitude results are covered in Section 3.3.1 and Section 3.3.2, respectively, and those for radio and FIR stacking are detailed in Section 3.3.3 and Section 3.3.4, respectively.

3.3.1 Lyman-alpha

Lyα is emitted from ionised gas around star forming regions and AGN, and therefore traces O and B type stars ($\mathrm{M} \geq 10$ $\mathrm{M_\odot}$) with lifetimes of the order of $\sim$ Myrs. Hence, Lyα can give an excellent indication of the instantaneous SFR in a galaxy. However, a significant fraction of Lyα photons are scattered by the interstellar medium (ISM) leading to kpc-long random walks. As a consequence, there is a very high probability that Lyα photons are absorbed by dust particles and destroyed. This leads to the amount of photons we receive being lower than expected, resulting in lower calculated SFRs (Sobral & Matthee 2019).

Through the process of excluding AGN from the catalogue we found that, in our original sample of 3908 sources,

109 emitted strongly in the X-ray, and 62 emitted strongly in the radio. 142 sources emitted strongly in either or both, leaving 3766 sources available for us to investigate Lyα SFRs. Some of the average properties of this sample are presented in Table 3.

We follow Sobral & Matthee (2019) for the dust corrected SFR, in $M_{\odot}yr^{-1}$, based on Kennicutt (1998) and a Salpeter IMF ($0.1 - 100$ $M_{\odot}$) using Equation 6:

$$SFR_{Ly\alpha} = \frac{L_{Ly\alpha} \times 7.9 \times 10^{-42}}{(1 - f_{esc,LyC})(0.042EW_0)} (\pm 15\%) \quad (6)$$

where $L_{Ly\alpha}$ is the Lyα luminosity, EW_0 is the rest frame equivalent width, and $f_{esc,LyC}$ is the Lyman continuum escape fraction. We assume that $f_{esc,LyC} = 0$.

Once we had obtained individual SFRs for each source, we averaged the SFR of all sources in each redshift bin, in order to obtain an average SFR per redshift bin, with associated standard deviations on both the SFR and the redshift. The redshift bin allocations are outlined in Table 3.

We found that the average $SFR_{Ly\alpha}$ was 16.5 ± 0.3 $M_{\odot}yr^{-1}$, with a standard deviation of 21.0 $M_{\odot}yr^{-1}$ over the 3766 sources.

3.3.2 Rest-frame Ultraviolet

Younger and more massive stars (O and B type) emit strongly in the UV, and in a similar fashion to Lyα, we can use rest frame UV magnitudes to calculate instantaneous SFRs. Sobral et al. (2018b) gives the UV SFR in $M_{\odot}yr^{-1}$ using Equation 7:

$$SFR_{UV} = (1.4 \times 10^{-28}) 4\pi \times 9.521 \times 10^{38} \times 10^{-0.4(M_{UV}+48.6)} \quad (7)$$

where the absolute UV magnitude M_{UV} is calculated using Equation 8:

$$M_{UV} = m_{UV} - 5\log_{10}\left(\frac{d_L}{10\,pc}\right) + 2.5\log_{10}(1+z) \quad (8)$$

where we use the rest frame B band from SED fitting as in Ilbert et al. (2009) for m_{UV} , and d_L is the luminosity distance. Note that we use the B band as in Ilbert et al. (2009) for m_{UV} across the whole redshift range, but this method will become less accurate as we move to higher redshifts, as the wavelength we observe the rest frame to be at increases. Future work to improve on this would be to use different bands for different redshifts.

In the same way as for Lyα, we averaged individual SFRs over the redshift bins outlined in Table 3. We found that the average SFR_{UV} to be 4.55 $M_{\odot}yr^{-1}$ with a standard deviation of 4.73 $M_{\odot}yr^{-1}$ across 3766 sources.

3.3.3 Radio

We made use of data from radio stacking; a technique in which data from many individual objects is combined in order to determine the average properties of sources that are too faint to be detected individually. Stacked values of certain average galaxy properties, such as flux densities, from Calhau (2019) were used for given redshifts. The stacking process results in these values being upper limits, rather than direct measurements. Once the determined AGN sources were removed from the data set, leaving only non-detections, the luminosity distance d_L of each redshift bin was calculated from the median z value. This was done using a calculation following Hogg (1999), with the Hubble constant H_0 and energy density distributions stated in Section 1. The obtained d_L values (used for both radio and FIR stacking methods) are presented in Table 3.

Following Calhau (2019), the upper limits of flux densities for 1.4 GHz were used to determine the upper limits of luminosities $L_{1.4GHz}$ in WHz^{-1} using Equation 9,

$$L_{1.4GHz} = 4\pi d_L^2 S_{1.4GHz} \times 10^{-33}(1+z)^{\alpha-1} \quad (9)$$

where d_L is the luminosity distance in cm, $S_{1.4GHz}$ is the flux density at 1.4 GHz in mJy, z is redshift and α is the the radio spectral index, assumed to be 0.8 (Delhaize et al. 2017). Often, a similar process is also carried out for frequency 3.0 GHz as this probes deeper still than 1.4 GHz. In this work, however, we focus just on 1.4 GHz radio luminosity.
The upper limits of $SFR_{1.4GHz}$, in $M_{\odot}yr^{-1}$, could then be calculated from the obtained luminosities following Calhau (2019) (submitted) using Equation 10 (Karim et al. 2011).

$$SFR_{1.4GHz} = 3.18 \times 10^{-22} L_{1.4GHz} \quad (10)$$

These values are for a Chabrier IMF and as such were later scaled by a factor of $\frac{7.9}{4.4}$ to a Salpeter IMF for comparison with the values obtained through the methods outlined in Section 3.3.1. The SFR upper limits calculated (103 $M_{\odot}yr^{-1} \lesssim SFR_{1.4GHz} \lesssim 2440$ $M_{\odot}yr^{-1}$) are discussed in further detail in Section 6.2.

3.3.4 Far Infrared

Similarly to the process carried out for radio stacking, SFR upper limits were determined using stacked FIR data from Calhau (2019). The original flux density limits (actually $3\times$ noise, rather than "data points") in the catalogue were in units of Jy and first had to be converted into flux densities per wavelength. Once this was done, the flux densities for each redshift were plotted against rest frame wavelength, λ_0, which were calculated from the observed wavelengths, λ_{obs}, provided in the data catalogue using the relation given by Equation 11.

$$\lambda_0 = \frac{\lambda_{obs}}{(1+z)} \quad (11)$$

To obtain the upper limits of total flux across the FIR wavelength range, the flux densities were also converted into rest frame values by multiplying them by a factor of $(1+z)$ before integrating. As all that was available was the given flux density values, the functional form of this plot was unknown. This was approximated as a Planck black body curve, shown by Equation 12, which was manually modified in each case to best fit the upper limit values. This approximation can be made as FIR emissions are due to the absorption of UV radiation by dust in HII regions and its subsequent re-emission in the far infrared range after the de-excitation of the heated dust particles. This black-body-like radiation corresponds to a temperature in the range $\sim 20 - 80$ K.

The method used is a simplified form of spectral energy distribution (SED) fitting and an example for $z \sim 3.3$ is shown by Figure 2.

Filter	λ_c [Å]	Average $L_{Ly\alpha}$ [$\log_{10}$(erg s^{-1})]	Average Mass [$\log_{10}(M_\odot)$]	Bin	No. Sources	$z_{1.4GHz,FIR}$	d_L [10^4 Mpc]
NB392	3920	42.6	9.22	$z \sim 2.2$	146	2.51	2.05
IA427	4270	42.7	9.09	$z \sim 2.5$	714	2.82	2.35
IA464	4640	42.9	9.16	$z \sim 3.1$	293	2.98	2.52
IA484	4840	42.9	9.12	$z \sim 3.1$	673	3.15	2.70
NB501	5010	43.0	9.38	$z \sim 3.1$	39	3.33	2.88
IA505	5050	42.9	9.28	$z \sim 3.1$	467	3.72	3.29
IA527	5270	42.9	9.27	$z \sim 3.1$	626	4.13	3.72
IA574	5740	43.0	9.38	$z \sim 3.9$	96	4.58	4.21
IA624	6240	43.0	9.22	$z \sim 3.9$	141	4.83	4.48
IA679	6790	43.3	9.42	$z \sim 4.7$	76	5.07	4.74
IA709	7090	43.19	9.29	$z \sim 4.7$	80	5.31	5.00
NB711	7110	42.8	9.10	$z \sim 4.7$	78	5.71	5.45
IA738	7380	43.27	9.46	$z \sim 5.4$	78	–	–
IA767	7670	43.4	9.48	$z \sim 5.4$	32	–	–
NB816	8160	42.9	9.40	$z \sim 5.4$	192	–	–
IA827	8270	43.5	9.69	$z \sim 5.4$	35	–	–

Table 3. The redshift bin allocations for our sample (Sobral et al. 2018a). $z_{1.4GHz,FIR}$ are the median redshift values from the stacked radio and FIR data (Calhau et al. in prep.) detailed in Section 3.3.3 and Section 3.3.4 and correspond to luminosity distances d_L. The – is due to lack of values.

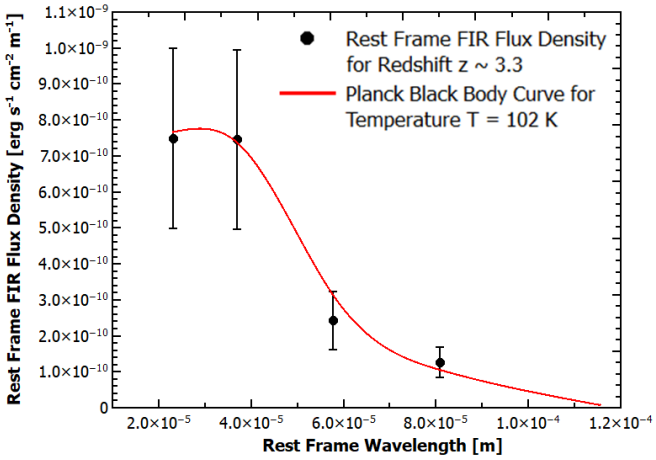

Figure 2. The rest frame FIR flux density (3 × noise) against rest frame wavelength for $z \sim 3.3$. The red curve is a Planck black body curve, given by Equation 12 for temperature T = 102 K, with its amplitude modified to best fit the flux density limits. This temperature is higher than might be expected (on the order of $\sim$ 10 K) but this is due to the limitations of the black body approximation. The value of T was chosen in each case to fit the values, rather than directly corresponding to the properties of the emission. Another potential SED fitting approach to improve on this is to make the "greybody" approximation discussed in Casey (2012) but not applied in this work.

This curve was then integrated with respect to wavelength, providing the upper limit of total flux for the given rest frame wavelength range.

$$B_\lambda(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{\exp(\frac{hc}{\lambda k_B T}) - 1} \quad (12)$$

B_λ is the spectral radiance for wavelength λ, T is the temperature, h is Planck's constant, and k_B is Boltzmann's constant. From the flux obtained through the integration of this function, luminosities L_{FIR} could then be calculated in units of erg s^{-1} using Equation 13,

$$L_{FIR} = 4\pi d_L^2 f_{FIR} \quad (13)$$

where d_L is the luminosity distance in cm and f_{FIR} is the flux in erg s^{-1}cm^{-2}. SFR_{FIR} values, in $M_\odot yr^{-1}$, could then be calculated from the luminosities obtained using the relation given in Kennicutt (1998), as displayed by Equation 14 .

$$SFR_{FIR} = 4.5 \times 10^{-44} L_{FIR} \quad (14)$$

The SFR upper limits determined from FIR stacking data (68.4 $M_\odot yr^{-1} \lesssim SFR_{1.4GHz} \lesssim 1820\ M_\odot yr^{-1}$) are discussed in Section 6.2.

3.3.5 Star Formation Rate Density

Star formation rate densities (SFRDs) are essentially the amount of stars forming per unit co-moving volume. It is important to note that the SFRDs in this paper were calculated differently to the SFRs. To determine the Lyα SFRD in $M_\odot yr^{-1}Mpc^{-3}$, we follow Sobral et al. (2018a) using Equation 15:

$$SFRD_{Ly\alpha} = \frac{\rho_{Ly\alpha} \times 7.9 \times 10^{-42}}{(1 - f_{esc,LyC}) \times 0.042 EW_0} \quad (15)$$

where $\rho_{Ly\alpha}$ is the Lyα luminosity density. We follow Sobral et al. (2018a) in deriving $\rho_{Ly\alpha}$ by integrating the Lyα luminosity function for each redshift slice in the SC4K sample.

To determine the UV SFRD, we took the rest frame UV SFRs calculated previously as in Section 3.3.2, and produced histograms of the weighted count of each galaxy with a certain SFR for each redshift bin. The weighted count n was calculated by dividing the number of sources N by the co-moving volume of the redshift slice V multiplied by the width of the SFR bin (ΔSFR), as shown in Equation 16.

$$n = \frac{N}{V \times \Delta log(SFR)} \quad (16)$$

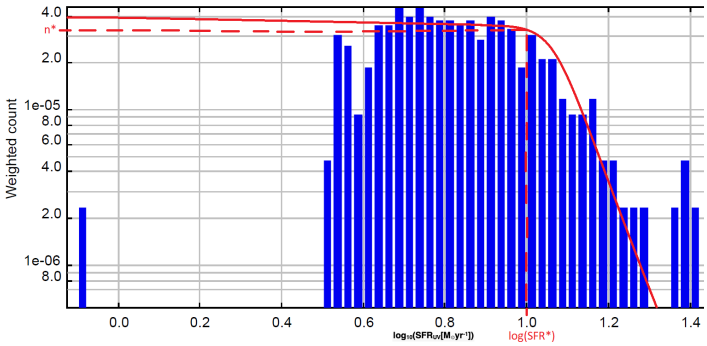

Figure 3. Determining n* and SFR*, which when combined in Equation 17 can be used to calculate $\mathrm{SFRD_{UV}}$.

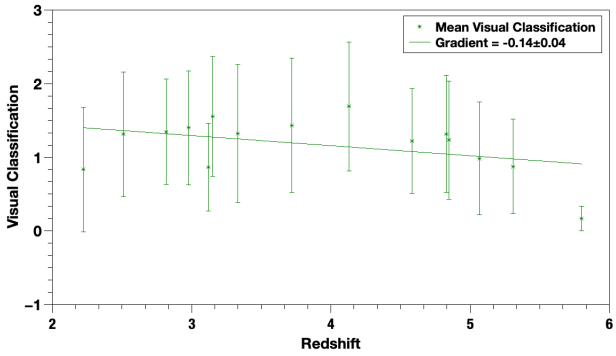

Figure 4. Mean visual classification per redshift slice against redshift, with standard deviation errors in which the majority of data points lie in the 1 < VC Mean < 2 range (compact to disky).

Visual Classification	VC Mean	Percentage of Sources (%)
Point-like	$-0.5 <$ VC Mean < 0.5	19.7 ± 1.7
Compact	$0.5 <$ VC Mean < 1.5	39 ± 3
Disky	$1.5 <$ VC Mean < 2.5	30 ± 2
Irregular	$2.5 <$ VC Mean < 3	11.3 ± 1.3

Table 4. The distribution of the VC Mean Trusted data set (Table 1, Section 3.1) over the classification scale described in Section 3.1.1.

An approximate Schechter function was fitted manually to each histogram, where the 'typical' SFR (SFR*) and 'typical' number density (n*) could be approximated. Then, using Equation 17, the UV SFRD was calculated. The errors in $\mathrm{SFRD_{UV}}$, in Figure 15, are proportional to the confidence in which we determined n* and SFR* using the histograms. An example is shown in Figure 3.

$$\mathrm{SFRD_{UV}} = \int \Phi \, d\mathrm{SFRD} = \mathrm{SFR}^* \times \mathrm{n}^* \times \Gamma(\alpha + 2) \qquad (17)$$

Φ is the Schechter function and Γ is the error function. We use α = -1.8.

4 RESULTS

4.1 Morphologies

4.1.1 Visual Classifications

Of the 788 galaxies in the reduced data set, most fell into the compact and disky ranges (see Table 4 for percentage distributions) while the average value across all redshifts is 1.3 ± 0.8. There is some fluctuation of visual classification mean (VC Mean) per redshift about this value which can be seen in Figure 4, with VC Mean per redshift starting at 0.17 ± 0.17 at redshift 5.8, climbing to 1.7 ± 0.9 at redshift 4.1 before declining to 0.8 ± 0.8 at redshift 2.2. See Table 5 for more detail.

4.1.2 GALFIT Analysis of SC4K LAEs

As can be seen in Figure 5, there is a tentative increase in all measurements of galactic radius with age (increase in $r_{20} = 0.05 \pm 0.02$ kpc per unit redshift, increase in $r_{50} = 0.09 \pm 0.04$ kpc per unit redshift, increase in $r_{80} = 0.16 \pm 0.08$ kpc per unit redshift and increase in $r_e = 0.18 \pm 0.07$ kpc per unit redshift) with the 50% to 80% light radii range increasing ~ 1.75 times faster than that of the 20% to 50% range. Our results agree well with those of Paulino-Afonso et al. (2018), Pirzkal et al. (2007) and Malhotra et al. (2012) in that, while there is a minor increase of radii with cosmic time, this is $\sim 2\sigma$ and is around an average radius of ~ 1 kpc. The Sérsic index has an overall mean of 1.9 ± 2.2 (this large error is due to the wide range of n values, $0 < n < 10$, despite 50% of all n values falling between 0 and 1.5) this is consistent with $z \sim 2-6$ SC4K LAEs being on average compact disks (which have $n \sim 2$) from Section 4.1.1. Figure 6 shows a similar pattern to Figure 4, if to a lesser extent (and within error bars), of starting low (0.9 ± 0.7 at redshift 5.8), increasing so that most mean n per redshift slice value lie above the fitted line and then decreasing again (to 1.4 ± 1.8 at redshift 2.2). The overall mean compactness is 2.6 ± 0.4 and this stays consistent over time (with a rate of change of -0.00 ± 0.07 per unit redshift) within the standard deviation. This result, and that of the Sérsic index data are consistent with results from Paulino-Afonso et al. (2018) which found a mean compactness C$\sim$ 2.7, and Sérsic index $n \lesssim 2$.

4.1.3 Stellar Masses

Figure 7 shows a tentative decrease in mass with age at a rate of $10^{0.09 \pm 0.05}$ $\mathrm{M_\odot}$ per unit redshift. Also shown in Figure 7 are the $\mathrm{M_{stellar}}$ values from Pirzkal et al. (2007) and Shapley et al. (2005), our data does not fit well with most of the Shapley et al. (2005) data and all but one data point from Pirzkal et al. (2007) which predict a gradient opposite to the one produced by our data. However the decrease in mass with age shown by our data is within the error bars which are large due to the large range of masses contained within the means.

Redshift z	VC Mean	r_{20} (kpc)	r_{50} (kpc)	r_{80} (kpc)	C	r_e (kpc)	n	$M_{stellar}$ ($\log_{10}(M_\odot)$)
2.22	0.8 ± 0.8	0.5 ± 0.3	1.0 ± 0.6	1.6 ± 1.0	2.6 ± 0.4	1.7 ± 1.4	1.4 ± 1.8	9.2 ± 0.6
2.51	1.3 ± 0.8	0.42 ± 0.13	0.8 ± 0.3	1.4 ± 0.5	2.6 ± 0.4	1.0 ± 0.6	2 ± 2	9.1 ± 0.6
2.82	1.3 ± 0.7	0.42 ± 0.13	0.8 ± 0.3	1.5 ± 0.6	2.7 ± 0.4	1.1 ± 0.7	2 ± 2	9.2 ± 0.5
2.98	1.4 ± 0.8	0.39 ± 0.13	0.8 ± 0.3	1.4 ± 0.6	2.7 ± 0.4	1.0 ± 0.6	2 ± 2	9.1 ± 0.6
3.12	0.9 ± 0.6	0.34 ± 0.13	0.7 ± 0.3	1.2 ± 0.5	2.6 ± 0.4	0.9 ± 0.6	2 ± 2	9.4 ± 0.5
3.15	1.6 ± 0.8	0.41 ± 0.14	0.8 ± 0.3	1.4 ± 0.6	2.6 ± 0.4	1.0 ± 0.6	2 ± 2	9.3 ± 0.6
3.33	1.3 ± 0.9	0.40 ± 0.12	0.8 ± 0.3	1.4 ± 0.5	2.6 ± 0.4	1.0 ± 0.6	2 ± 2	9.3 ± 0.6
3.72	1.4 ± 0.9	0.39 ± 0.13	0.8 ± 0.3	1.4 ± 0.5	2.7 ± 0.4	1.0 ± 0.6	3 ± 3	9.4 ± 0.5
4.13	1.7 ± 0.9	0.35 ± 0.09	0.7 ± 0.2	1.2 ± 0.4	2.6 ± 0.4	0.9 ± 0.6	2.0 ± 1.8	9.2 ± 0.6
4.58	1.2 ± 0.7	0.35 ± 0.11	0.7 ± 0.3	1.3 ± 0.5	2.7 ± 0.3	0.9 ± 0.6	2 ± 2	9.4 ± 0.5
4.83	1.3 ± 0.8	0.36 ± 0.10	0.69 ± 0.19	1.3 ± 0.4	2.7 ± 0.4	0.9 ± 0.5	2 ± 2	9.3 ± 0.5
4.85	1.2 ± 0.8	0.31 ± 0.10	0.6 ± 0.2	1.1 ± 0.5	2.6 ± 0.4	0.8 ± 0.5	1 ± 3	9.1 ± 0.5
5.07	1.0 ± 0.8	0.32 ± 0.11	0.7 ± 0.2	1.2 ± 0.4	2.7 ± 0.4	0.9 ± 0.5	3 ± 3	9.5 ± 0.4
5.31	0.9 ± 0.6	0.32 ± 0.12	0.6 ± 0.2	1.2 ± 0.5	2.8 ± 0.5	0.7 ± 0.5	2.1 ± 1.5	9.5 ± 0.2
5.71	–	0.24 ± 0.10	0.5 ± 0.2	0.8 ± 0.4	2.5 ± 0.5	0.7 ± 0.5	2 ± 2	9.4 ± 0.4
5.80	0.16 ± 0.16	0.25 ± 0.09	0.5 ± 0.2	0.9 ± 0.4	2.5 ± 0.4	0.5 ± 0.3	0.9 ± 0.7	9.7 ± 0.3

Table 5. The mean values and standard deviation per redshift slice of the reduced data sets (see Table 1) considered in this study. VC Mean is the mean visual classification (0 being point-like, 1 is compact, 2 is disky and 3 is irregular, see Section 4.1.1 for more details on these results). r_{20}, r_{50}, r_{80} and r_e are the 20%, 50%, 80% and fitted half light radii (the radii at which 20%, 50%, 80% and half the light falls within said radius), and C and n are the compactness and the Sérsic index respectively (see Section 4.1.2 for more on these results). $M_{stellar}$ is the logged stellar mass of the LAE (see Section 4.1.3 for more on these results). The – in the VC Mean column is due to a lack of any sources for this redshift slice, as in Table 1.

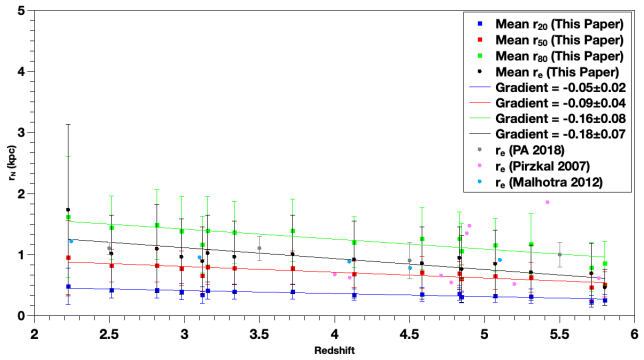

Figure 5. The sizes of our LAEs as a function of redshift for different radii (r_N) and their standard deviations, where dark blue is r_{20}, red is r_{50}, green is r_{80} and black is r_e which are the radii at which 20%, 50%, 80% and half the light falls within said radius, respectively (see Section 3.1.2). Also shown are r_e data from Paulino-Afonso et al. (2018) in grey, Pirzkal et al. (2007) in pink, and Malhotra et al. (2012) in light blue.

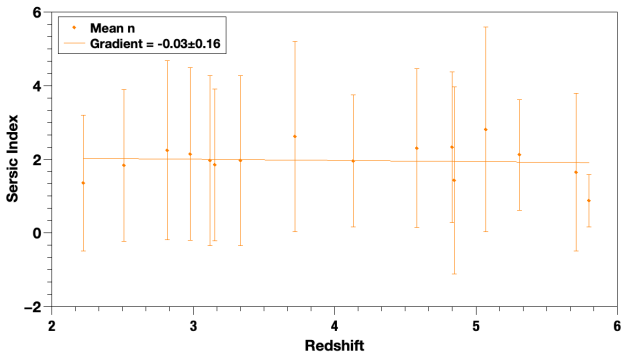

Figure 6. Mean Sérsic index and standard deviation per redshift slice against redshift showing no evolution which is consistent with our LAEs on average compact disks ($n \sim 2$) from $z \sim 2-6$.

4.2 AGN

From the *Chandra* and VLA data, 303 AGN sources in the SC4K catalogue were identified, 240 of which are detected at X-ray wavelengths and 119 are detected in the 1.4 GHz and 3.0 GHz radio bands combined. For all sources detected in the X-ray wavelength the luminosities calculated where found in a range between $\sim 10^{42}$ erg s^{-1} and $\sim 10^{45}$ erg s^{-1}. These luminosities correspond to a range of black hole accretion rates (BHAR) from ~ 0.03 $M_\odot$ yr^{-1} to ~ 3.3 $M_\odot$ yr^{-1}, which is consistent with the range of BHARs expected across this redshift range (e.g. Calhau 2019; Shields 1999). Two AGN were also found to have much larger BHARs of (6.9 ± 2.3) $M_\odot$ yr^{-1} and (10.0 ± 0.1) $M_\odot$ yr^{-1}, indicating highly active AGN in our sample.

The fractions of AGN in the SC4K catalogue across the range of Lyα luminosities and across the redshift range of the catalogue are plotted (see Section 4.2.1 and Section 4.2.2 respectively) and the trend in the relation between BHAR and the morphologies of the AGN in our catalogue (shown in Section 4.2.3) is explored, in order to investigate the luminosity, size and evolutionary trends of the AGN identified in the SC4K catalogue.

4.2.1 AGN Activity Dependence on Lyman-alpha Luminosity

The AGN identified in the SC4K catalogue were found to have Lyα luminosities between $L_{Ly\alpha} \sim 10^{44.5}$ erg s^{-1} and

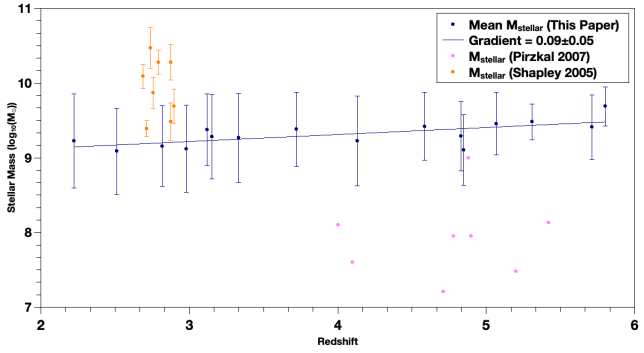

Figure 7. Graph of mean stellar mass and standard deviations per redshift slice against redshift. Where navy is $M_{stellar}$ from this paper, pink is $M_{stellar}$ from Pirzkal et al. (2007), and orange is $M_{stellar}$ from Shapley et al. (2005).

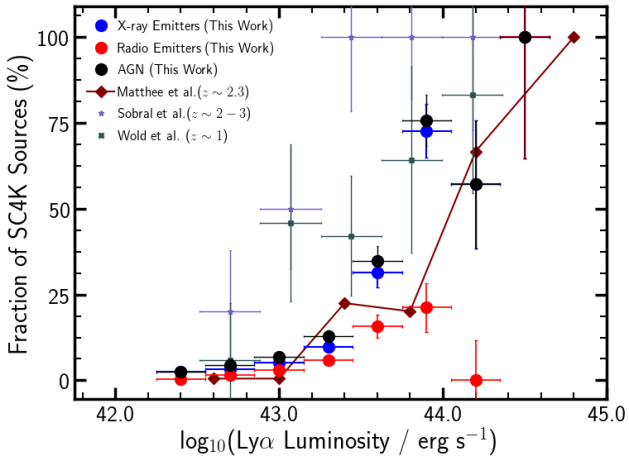

Figure 8. The faction of X-ray AGN (blue), radio AGN (red) and the combined AGN fraction (black) in the SC4K catalogue, in luminosity $L_{Ly\alpha}$ bins of $10^{0.3}$ erg s^{-1}, showing the rapid increase in the fraction of AGN as the luminosity increases. Similar results have been found in previous studies, Matthee et al. (2017) (maroon) is show here with our data. For comparison we also show the AGN fractions found using spectroscopic data across redshifts of $z \sim 1$ (grey) and $z \sim 2-3$ (lilac) from Sobral et al. (2018b); Wold et al. (2014) respectively.

$L_{Ly\alpha} \sim 10^{42.4}$ erg s^{-1} which is consistent with the expected Lyα luminosities of our AGN (e.g. Calhau 2019).

Plotting the fraction of all active galaxies in our catalogue, those which are X-ray or radio emitters and the combined fraction, against the Lyα luminosity (see Figure 8) shows that the fraction of all galaxies which are AGN increases rapidly with the luminosity, with only a small fraction at $\sim 10^{42}$, increasing almost exponentially as luminosity increases up to $\sim 10^{45}$. This trend is in agreement with that found in Matthee et al. (2017), shown on Figure 8, and shows a similar AGN fraction to that found with spectroscopic data taken in Sobral et al. (2018b) and Wold et al. (2014).

Whilst there is a general increase in the fraction of sources emitting in X-ray or Radio, which flows the trends found in previous studies, it can be seen that the fraction of sources emitting in the Radio range is consistently lower than that of the fraction emitting X-rays. This difference becomes more pronounced at higher luminosities.

At high luminosities the sample size is small which may cause the results to be skewed due to sample bias, which causes the uncertainties on the AGN fraction to increase with the Lyα luminosity of the AGN.

4.2.2 Redshift Dependence of AGN Fraction

Plotting the fraction of AGN in our catalogue at different redshift bins shows that the fraction of AGN remains relatively low, remaining below 15% of the total sources, across the range of the catalogue with only slight fluctuations (see Figure 9), following a similar trend to that found in previous studies such as Calhau (2019) and Lehmer et al. (2013a).

As shown in Figure 9, the fraction of radio sources shows a gradual increase as the redshift value decreases, peaking slightly at $z \sim 3-4$ before decreasing again from $z \sim 3$ to $z \sim 2$. The fraction of X-ray AGN in our catalogue shows a similar yet less distinct trend with redshift, whilst peaking at $z \sim 3$ and $z \sim 5.5$. From the fraction of AGN in the redshift ranges of $z \sim 2-4$ and $z \sim 4-6$ the overall decrease in the source fraction with redshift can been seen, with the percentage fractions of X-ray, radio and the combined AGN fraction given in Table 6.

LAEs	$z \sim 2-4$ fraction [%]	$z \sim 4-6$ fraction [%]
Total in Catalogue	81.6 ± 0.6	18.4 ± 0.6
X-ray emitters	6.9 ± 0.4	2.8 ± 0.6
Radio emitters	3.4 ± 0.3	1.3 ± 0.4
AGN (X-ray+Radio)	8.7 ± 0.5	3.6 ± 0.0

Table 6. The fraction of X-ray AGN, radio AGN and the combined total AGN fraction in the redshift ranges of $z \sim 2-4$ and $z \sim 4-6$ showing the overall decrease in AGN fraction with the increase in redshift. The fraction of all sources in each redshift range is also shown.

Thus the AGN fraction in each redshift bin decreases with an increase in redshift, peaking at cosmic noon ($z \sim 2-3$), as has been observed in previous studies (e.g. Calhau 2019). The peak observed at the redshift of $z \sim 5$ is likely due to sample bias as only brighter LAEs are detected at high redshifts which are found to be mostly AGN (see Section 4.2.1), and the sample size at higher redshifts is small.

4.2.3 Morphological Trends in SC4K AGN

From the morphology calculations the light radii of the galaxies in our catalogue have been found (detailed in Section 3.1). This allows the morphological trends of the AGN in the range of $z \sim 2-6$ to be investigated. From our 303 AGN sources, only 198 had percentage light radii calculated and only 53 have Sérsic indices which could be calculated to a reasonable degree of accuracy (see Table 1).

Plotting the BHAR of those AGN, for which the measured BHAR values could be calculated, against the 20%,

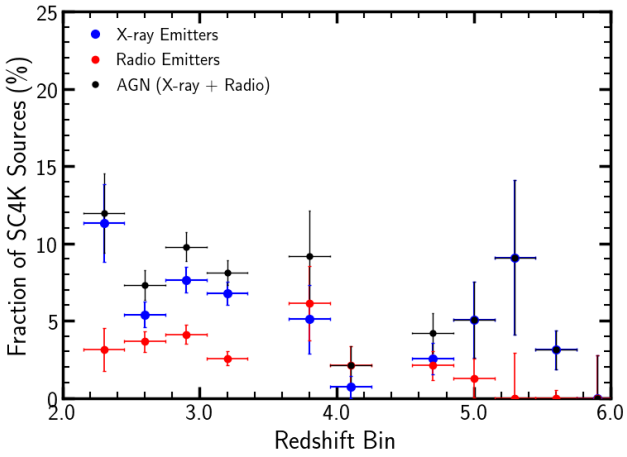

Figure 9. The fraction of X-ray AGN (blue), radio AGN (red) and the combined AGN fraction (black) in the SC4K catalogue, in redshift bins of 0.3 (indicated by the errors in redshift), showing two slight peaks at $z \sim 3$ and $z \sim 5$ and an overall gradual downward trend in the AGN fraction. The binomial errors in the fraction of sources detected are shown.

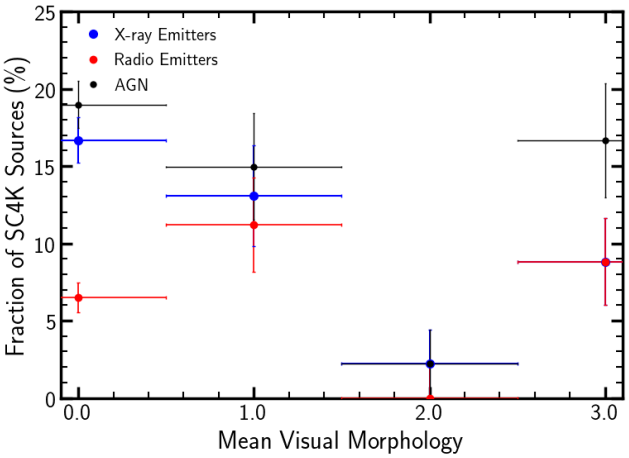

Figure 10. The fraction of X-ray (blue), radio (red) and the combined AGN fraction (black) with visual morphologies of 0, 1, 2 and 3, corresponding to point-like, compact, disky and irregular morphologies respectively, showing a peak in the fraction of AGN at point-like and compact morphologies and again for irregular morphologies. This is likely due to the compact nature of the bright centres of Active Galaxies and the radio jets that can be detected from AGN.

50% and 80% light radii, shows the spread of AGN sizes, see Figure 11, with most of the light being emitted about the central region within which is the supermassive black hole. 80% of the light emitted by the whole active galaxy is within ~ 4.5 kpc of the centre, with 50% of the light emitted within ~ 3.0 kpc and 20% emitted within ~ 2.0 kpc of the centre of the AGN. From the spread of the data points it can be seen that the most active sources, for which the percentage light radii have been calculated, typically have smaller light radii at each percentage, with the percentage light radius decreasing with the increase in BHAR. Plotting the half light radii calculated using the Sérsic profile (see Section 3.1.2) of our AGN, against the BHAR values shows that the half light radius decreases rapidly, from a radius of ~ 4.75 kpc to a radius of ~ 0.1 kpc, with increasing BHAR. Thus most light emitted by active galaxies is produced by the supermassive black hole in the centre of these galaxies, with the more active AGN being more luminous in the central region the galaxy.

From the visually classified morphologies (see Section 3.1.1) the fraction of AGN in our catalogue with the different visual morphologies shows that a greater percentage of AGN have visual morphologies of 0 and 1, with the X-ray and radio fractions peaking a visual morphologies of 0 and 1 respectively (see Figure 10). A slight peak in the AGN fraction also occurs for a visual morphology of 3. Thus AGN are typically more point-like compact sources with some appearing as more irregular sources, which is consistent with what we expect. This concurs with the half light radii, from the Sérsic profiles, and the percentage light radii results, indicating AGN typically appear as very bright point-like sources, with little of the surrounding galaxy visible due to the high luminosity of the supermassive black hole itself.

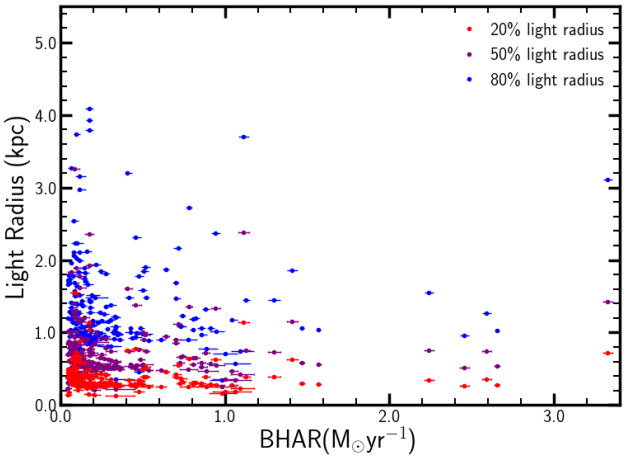

Figure 11. The 20% (red), 50% (Purple) and 80% (blue) light radii, against the BHAR with AGN of higher BHAR having smaller percentage light radii. From the scatter the range of light radii at 20%, 50% and 80% can be seen as $\sim 0.1-0.5$ kpc, $\sim 0.5-1.0$ kpc and $\sim 1.2-3.0$ kpc respectively.

4.3 SFR

Figure 12 (bottom right panel) shows that SFR is generally greater at higher redshifts, except at redshifts $z \gtrsim 5$. Between redshifts $4.8-5$, we no longer see SFR increase with increasing z. Instead, the results show that a given SFR occurs at a higher stellar mass M. Across the stellar mass range, we see a local maximum in SFR for each redshift slice. Averaged across the redshift range, this local maximum is located at $M \sim 10^{9.3}$ $M_\odot$ for $z \sim 2.5$. However, we see that

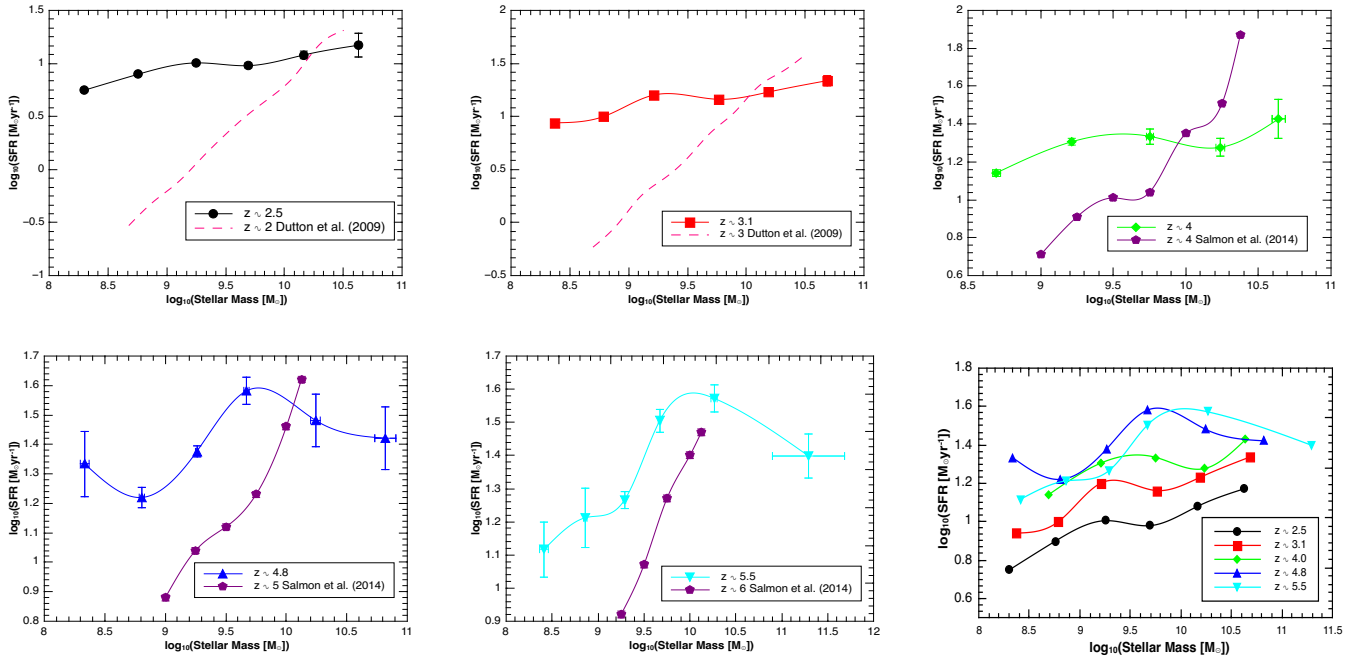

Figure 12. Lyα SFR evolution with stellar mass for varying redshift bins. Top left: $z \sim 2.5$; top middle: $z \sim 3.1$; top right: $z \sim 4.0$; bottom left: $z \sim 4.8$; bottom middle: $z \sim 5.5$; bottom right: combined results for all redshifts. Results from Dutton et al. (2010) and Salmon et al. (2015) are plotted for redshifts $z \sim 2.5 - 3.1$ and for redshifts $z \sim 4.0 - 5.5$, respectively. We find that in general SFR increases with stellar mass, and that SFRs are generally higher at higher redshifts. See Section 6.2 for further analysis.

this maximum appears to progressively shift to higher M with increasing z, peaking at M $\sim 10^{10}$ M$_\odot$ for $z \sim 5.5$. The amplitude of these maxima are also seen to increase with increasing redshift. At $z \sim 2.5$, the peak is at SFR $\sim 10^{1.0}$ M$_\odot$yr^{-1}, but is located at SFR $\sim 10^{1.6}$ M$_\odot$yr^{-1} for $z \sim 5.5$. However, the range of SFRs seems fairly consistent across the redshift range. At $z \sim 2.5$, the range of SFRs is $\sim 10^{0.4}$ M$_\odot$yr^{-1}, but at $z \sim 5.5$ the range is $\sim 10^{0.5}$ M$_\odot$yr^{-1}.

From redshifts $z \sim 2.5 - 4$ we see that SFR increases up to the maximum, dips briefly, then continues to increase again after the local maximum. However, at higher redshifts ($z \sim 4 - 5.5$) we see SFR decrease after the 'bump'. This is most likely due to the 'dust wall' created by galaxies with large stellar masses and high SFRs. This affects the Lyα escape fraction, and hence directly affects the SFR we calculate. According to Figure 13, this 'dust wall' occurs at SFRs greater than $\sim 10^{1.6}$ M$_\odot$yr^{-1}, and stellar masses greater than $\sim 10^{10}$ M$_\odot$.

Figure 13 shows that, averaged across the redshift range $z \sim 2 - 6$, SFR increases with stellar mass, with further results on the evolution of SFR with redshift presented in Figure 14. In Figure 13 we also see more clearly that the local maxima appear around a stellar mass of $\sim 10^{9.3}$ M$_\odot$ for both Lyα and UV. Given this, and that the shape of the curves are practically the same, SFR$_{\mathrm{Ly}\alpha}$ and SFR$_{\mathrm{UV}}$ are consistent, but differ in their scaling. The scaling difference gives an indication of the dust content and the ionisation rate around the galaxies. This relationship is represented by the ratio SFR$_{\mathrm{Ly}\alpha}$/SFR$_{\mathrm{UV}}$, also shown in green in Figure 13. This ratio peaks at M $\sim 10^{10}$ M$_\odot$, which is also the point where we found the 'dust wall' to become significant.

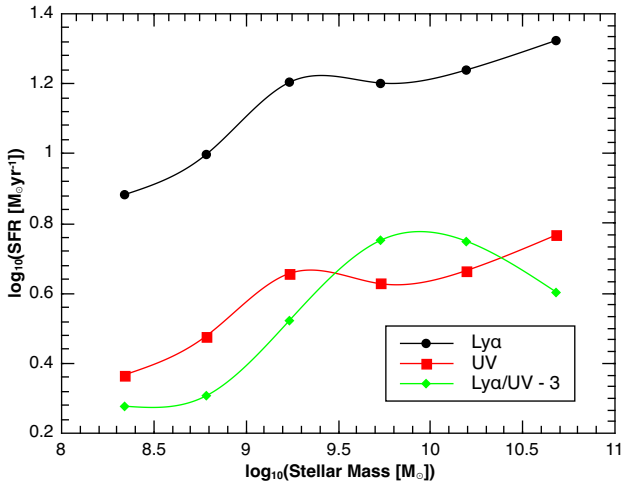

Figure 13. Averaged SFRs against stellar mass across the redshift range $z \sim 2 - 6$ calculated using Lyα luminosities and rest frame UV magnitudes. Green diamonds show the ratio Lyα/UV, which gives an indication of the dust content and ionisation fraction in galaxies. This ratio has been translated down by 3 in order to better compare against Lyα and UV.

Figure 14 shows that, across all our data, observational SFRs tend to increase with redshift. Specifically, we find that $\frac{d}{dz}\log_{10}(\frac{\mathrm{SFR}_{\mathrm{Ly}\alpha}}{\mathrm{M}_\odot \mathrm{yr}^{-1}}) = 0.140 \pm 0.004$, and $\frac{d}{dz}\log_{10}(\frac{\mathrm{SFR}_{\mathrm{UV}}}{\mathrm{M}_\odot \mathrm{yr}^{-1}}) =$

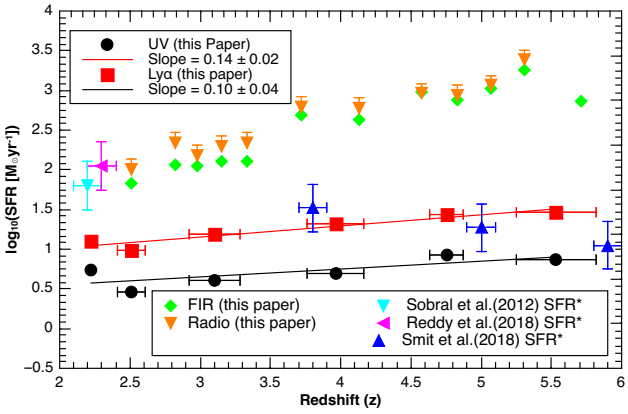

Figure 14. SFR evolution with redshift. Red squares: calculated using Lyα luminosities; black circles: calculated using rest frame UV magnitude; orange triangles: from 1.4 GHz radio stacking; green diamonds: FIR stacking. Results from Smit et al. (2012), including those derived from Sobral et al. (2012a) and Reddy et al. (2008), are plotted for comparison. Radio and FIR values are upper limits from non-detections (detailed in Section 6.2). We find that SFR increases with redshift across all tracers: Lyα, rest frame UV, FIR, and radio.

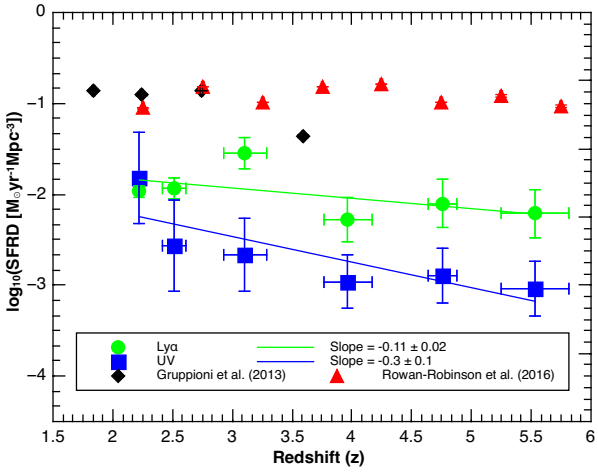

Figure 15. SFRD evolution with redshift. Green circles: calculated using Lyα luminosities; blue squares: calculated using UV magnitudes; black diamonds: Gruppioni et al. (2013); red triangles: Rowan-Robinson et al. (2016). We find that SFRD decreases with increasing redshift.

0.099 ± 0.004. This demonstrates that $\mathrm{SFR_{Ly\alpha}}$ diverges from $\mathrm{SFR_{UV}}$, which is due to the presence of dust and a varying ionisation rate. Radio data provides upper limits on the SFR, and it is therefore in line with expectation that they form the 'ceiling' values in Figure 14. As with Lyα and UV, $\mathrm{SFR_{1.4GHz}}$ increases with redshift. No line of best fit was plotted because theses results are from non-detections. FIR data are also upper limits, and combined with UV would roughly trace the SFR given by radio stacking, and therefore give the total SFR, as FIR traces obscured star formation and UV traces unobscured.

In Figure 14, the errors on redshift, $\mathrm{SFR_{Ly\alpha}}$ and $\mathrm{SFR_{UV}}$ are standard errors, calculated from the standard deviation. If error bars cannot be seen on $\mathrm{SFR_{Ly\alpha}}$ and $\mathrm{SFR_{UV}}$, it is because they are contained within the points themselves.

The errors on the $\mathrm{SFR_{FIR}}$ results were derived from uncertainties on the calculated $\mathrm{f_{FIR}}$ upper limits. These were determined by fitting and integrating two black body curves – the best and least acceptable fits – to the flux density upper limits (3 × noise from stacking) for each redshift. The difference between the $\mathrm{f_{FIR}}$ value obtained from each integration was taken to be the error on the flux limit in each case. The fractional error in each case was equated to the fractional error for $\mathrm{SFR_{FIR}}$ values to determine the SFR uncertainty. As the $\mathrm{SFR_{FIR}}$ are upper limits, lower limit error bars were neglected. As such, only the positive error was plotted, effectively giving the absolute upper limit of SFR. This is not visible on Figure 14 as the errors were negligible compared to the scale of the plot.

The errors on the radio SFR values were calculated by equating the fractional errors of $\mathrm{SFR_{1.4GHz}}$ and $\mathrm{S_{1.4GHz}}$. This was done as, from Equation 9 and Equation 10, it can be seen that $\mathrm{SFR_{1.4GHz}}$ is proportional to flux density $\mathrm{S_{1.4GHz}}$. As with the FIR SFR results, only the positive errors were included on the Figure 14, giving the true upper limits.

Figure 15 shows that SFRD decreases with increasing redshift in the range $z \sim 2.2 - 5.5$. We find $\frac{d}{dz}\log_{10}(\frac{\mathrm{SFRD_{Ly\alpha}}}{\mathrm{M_\odot yr^{-1}Mpc^{-3}}}) = -0.11 \pm 0.02$, and $\frac{d}{dz}\log_{10}(\frac{\mathrm{SFRD_{UV}}}{\mathrm{M_\odot yr^{-1}Mpc^{-3}}}) = -0.3 \pm 0.1$. In a similar fashion to Figure 14, $\mathrm{SFRD_{Ly\alpha}}$ diverges from $\mathrm{SFRD_{UV}}$, again thought to be due to dust and the ionisation rate.

5 PROGENITOR ANALYSIS

When studying objects at high redshifts it is important to remember that the further away a galaxy is, the earlier in its evolutionary lifetime it appears. This means that, by looking at a wide range of redshifts, we can effectively watch the evolution of these young galaxies. In this section we attempt to study young LAE galaxies that will eventually evolve into one of 3 types of $z = 0$ galaxies, these are dwarf, Milky Way-like, and brightest cluster galaxies (BCGs). Their earlier forms are referred to as dwarf-like, Milky Way-like and cluster-like progenitor galaxies. In the study of the evolutionary path of these progenitors, and particularly those in the Milky Way-like category, we can learn much about the history of our own galaxy and those nearby to us.

5.1 Methodology

5.1.1 Identifying Progenitors using Luminosity

To find progenitors in the SC4K catalogue we used Khostovan et al. (2018) as reference, in which LAE dark matter halos masses are related to their luminosities. The dark matter evolutionary track of present day galaxies with halo masses $\sim 10^{11-14}$ $\mathrm{M_\odot}$ were plotted in their paper (in which dwarf galaxies, Milky Way-type galaxies and BCGs are classified as having $\sim 10^{11}$ $\mathrm{M_\odot}$, $\sim 10^{12}$ $\mathrm{M_\odot}$ and $> 10^{13}$ $\mathrm{M_\odot}$ respectively). Using data from said plot, values for the average dark matter

halo mass M_{halo} of dwarf-like, Milky Way-like and cluster-like galaxy progenitors were found at redshifts matching our redshift slices. This data was then used in Equation 18, to produce average values for the Lyα luminosity of progenitor galaxies at these same redshifts. In the case of cluster-like progenitors, their dark matter halo masses are greater than 10^{13} $M_\odot$ so instead of using the range of the luminosity errors as limits for choosing our progenitors, Equation 18 was used with the $M_{halo} = 10^{13}$ $M_\odot$ and $M_{halo} = 10^{14}$ $M_\odot$ lines from Khostovan et al. (2018) with the range of errors on the luminosities produced as the extended range of luminosities for cluster-like progenitors.

$$\log_{10}(L(z)) = \frac{1}{2.08^{+0.12}_{-0.12}}\left(\log_{10}\left(\frac{M_{halo}}{M_\odot}\right) - 12.19^{+0.06}_{-0.06}\right) + \log_{10}(L^\star(z)) \quad (18)$$

Where Equation 18 is derived (under the condition $L < L^\star$ as this is true for the majority of our Lyα luminosities, with the exception being $2 < z < 3.5$, $M_{halo} \sim 10^{14}$ $M_\odot$, cluster-like progenitors, for which $2.08^{+0.12}_{-0.12}$ is replaced by $0.63^{+0.12}_{-0.12}$) from Khostovan et al. (2018). $L(z)$ and M_{halo} are the average Lyα luminosity and dark matter halo mass of the progenitor in question, and $L^\star(z)$ is the Lyα characteristic line luminosity at redshift z. These values were then used to isolate which galaxies in our data set would later evolve into dwarf-like, Milky Way-like and cluster-like galaxies. No dwarf-like, a small number of Milky Way-like, and a much larger number of cluster-like progenitors were found (see Table 7). This is likely due to the selection bias present in high redshift detections as only the most massive of galaxies have luminosities high enough to be observed at such great distances. Thus the progenitors of cluster-like galaxies will be more highly represented than Milky Way-like progenitors as they are more massive and luminous. The same applies to Milky Way-like versus dwarf-like progenitors. The resulting sets of Milky Way-like and cluster-like progenitors were used to calculate mean values per redshift of the galactic properties listed in Section 3.1.1, Section 3.1.2 and Section 3.1.3 for both sets, along with standard deviation errors. Additionally, when discussing progenitor results the lookback time is used in place of redshift in order to better represent the progenitor galaxy's evolution with time.

5.1.2 Identifying Progenitors through SFR

To determine what SFR a galaxy with a specified mass needs in order to evolve into a galaxy with a mass similar to the Milky Way, we made the assumption that SFR is constant across cosmic time. With this assumption, we can easily integrate SFR with respect to time to determine how much mass is added to a galaxy over time.

$$M_{MWP} = M_{MW} - [SFR \times LBT] \quad (19)$$

M_{MWP} is the mass of the Milky Way progenitor galaxies, M_{MW} is the mass of the Milky Way, and LBT is the Lookback Time. We use $M_{MW} = 6.43 \times 10^{10}$ $M_\odot$ from McMillan (2011).

Figure 19 shows what mass a galaxy should have at the time we are observing them, for a given average SFR, in order to have a stellar mass equal to the Milky Way now.

5.2 Results

5.2.1 Progenitor Morphologies

For the set of Milky Way-like progenitors it can be seen from Figure 16 that all measurements of the progenitor's radii are increasing with age (increase in $r_{20} = 0.168 \pm 0.015$ pcMyr^{-1}, increase in $r_{50} = 0.38 \pm 0.03$ pcMyr^{-1}, increase in $r_{80} = 0.70 \pm 0.04$ pcMyr^{-1} and tentative increase in $r_e = 1.5 \pm 1.3$ pcMyr^{-1}) and, particularly for r_e, fit with results from the literature. As in Section 4.1.2, the 50% to 80% light radii range is increasing faster than that of the 20% to 50% range, if slightly slower (~ 1.5 times faster, rather than ~ 1.75 as before). As Figure 17 shows, the Milky Way-like progenitors have stellar masses which are increasing with age at a rate of $10^{0.1 \pm 1.0}$ $M_\odot$ Gyr^{-1}. This data doesn't fit as well as the r_e data but it does within standard deviation error. The progenitors' Sérsic index values do not fit well with literature values, however it does still fall within the range of its large error bars. There is only one Milky Way-like progenitor data point for VC mean which is 0.9 ± 1.1 which corresponds to a compact visual morphology, it does however have large errors due to a wide range of visual morphologies contained within this one mean.

Like those of the Milky Way-like progenitor radii, all measurements of cluster-like progenitor radii are increasing with age (increase in $r_{20} = 0.08 \pm 0.02$ pcMyr^{-1}, increase in $r_{50} = 0.16 \pm 0.04$ pcMyr^{-1}, increase in $r_{80} = 0.23 \pm 0.11$ pcMyr^{-1} and tentative increase in $r_e = 0.23 \pm 0.16$ pcMyr^{-1}). Unlike in Section 4.1.2 and with the Milky Way-like progenitors, the 50% to 80% light radii range are increasing at roughly the same rate as the 20% to 50% range. The results for cluster-like progenitor radii fit with the lowest results from Nelson et al. (2002), however, in order to fit with the majority of the rest of the results from Nelson et al. (2002), there would have to be a significant increase in rate of increase of radii (from a rate of increase in $r_e = 0.23 \pm 0.16$ pcMyr^{-1} to a, post lookback time = 10.5, rate of increase of ~ 4.5 pcMyr^{-1}, so a rate $\sim$ 20 times larger). Figure 18 shows that the cluster-like progenitors appear to have a greater rate of stellar mass loss than that of the full SC4K set with a tentative decrease of $10^{0.11 \pm 0.06}$ $M_\odot$ Gyr^{-1} which does not agree with data from Stott et al. (2010). There are still large errors on the points however, and they increase in size with age. The Sérsic index values for this set of progenitors fit well with the lower values of n from Nelson et al. (2002) and would only need a small increase in their rate of n increase with time to fit with the majority of the rest of the values from Nelson et al. (2002). However, they have very large error bars and have a very wide spread across a the range $0 < n < 2.5$. In the case of cluster-like galaxies there are more VC mean data points than for the Milky Way-like progenitors which shows almost no evolution with time (rate of VC mean decrease $= 0.1 \pm 0.3$ Gy^{-1}).

5.2.2 Progenitor SFR

Figure 19 shows that, for our sample, most early galaxies must have an average SFR $\lesssim 10^{0.73}$ $M_\odot$yr^{-1} over their lifetimes in order to have a stellar mass similar to that of the Milky Way today. However, given that our average SFR $= 10^{1.103 \pm 0.005}$ $M_\odot$yr^{-1}, galaxies in the early Universe

Data Set	Reduction Conditions	Total Number of Sources
Full SC4K	None	3908
DG Progenitor	$L(z) - \sigma(\text{down})_{L(z)} < L_{Ly\alpha}(z) < L(z) + \sigma(\text{up})_{L(z)}$	0
MW Progenitor	$L(z) - \sigma(\text{down})_{L(z)} < L_{Ly\alpha}(z) < L(z) + \sigma(\text{up})_{L(z)}$	62
CG Progenitor	$L_{min}(z) - \sigma(\text{down})_{L_{min}(z)} < L_{Ly\alpha}(z) < L_{max}(z) + \sigma(\text{up})_{L_{max}(z)}$	2984

Redshift z	SC4K Sources (%)	DG Progenitor Sources (%)	MW Progenitor Sources (%)	CG Progenitor Sources (%)
2.22	4.1 ± 0.3	–	61 ± 13	1.7 ± 0.2
2.51	19.0 ± 0.8	–	–	16.0 ± 0.8
2.82	8.0 ± 0.5	–	45 ± 10	10.4 ± 0.6
2.98	18.2 ± 0.7	–	–	22.8 ± 1.0
3.12	1.15 ± 0.17	–	–	1.4 ± 0.2
3.15	12.4 ± 0.6	–	–	15.6 ± 0.8
3.33	16.4 ± 0.7	–	–	20.4 ± 0.9
3.72	2.5 ± 0.3	–	–	2.9 ± 0.3
4.13	3.6 ± 0.3	–	–	0.03 ± 0.03
4.58	2.0 ± 0.2	–	–	–
4.83	2.1 ± 0.2	–	–	–
4.85	2.0 ± 0.2	–	–	2.0 ± 0.3
5.07	2.0 ± 0.2	–	–	1.8 ± 0.2
5.31	0.84 ± 0.15	–	–	0.10 ± 0.06
5.71	4.9 ± 0.4	–	1.6 ± 1.6	4.6 ± 0.4
5.80	0.90 ± 0.15	–	–	0.03 ± 0.03

Table 7. Table showing the reduction conditions used in Section 5.1.1 on the SC4K catalogue and the number of sources in the data sets produced, including the percentage of each data set's total number of sources in each redshift slice. Where DG, MW and CG Progenitors are dwarf-like, Milky Way-like and cluster-like progenitor galaxies, $L(z)$ and $\sigma(\text{down/up})_{L(z)}$ are the progenitor Lyα luminosity at redshift, z, and it's associated down/up error calculated using Equation 18 (similarly $L_{min/max}(z)$ and $\sigma(\text{down/up})_{L_{min/max}(z)}$are the min/max CG progenitor Lyα luminosities at redshift, z, and their associated down/up error see Section 5.1.1) and $L_{Ly\alpha}(z)$ is the Lyα luminosity of an SC4K galaxy at this same redshift. The – found in several columns is due to a lack of any sources for this redshift slice, as in Tables 1 and 5, large errors are caused by the comparatively large number of sources in a redshift slice in relation to the total for that set.

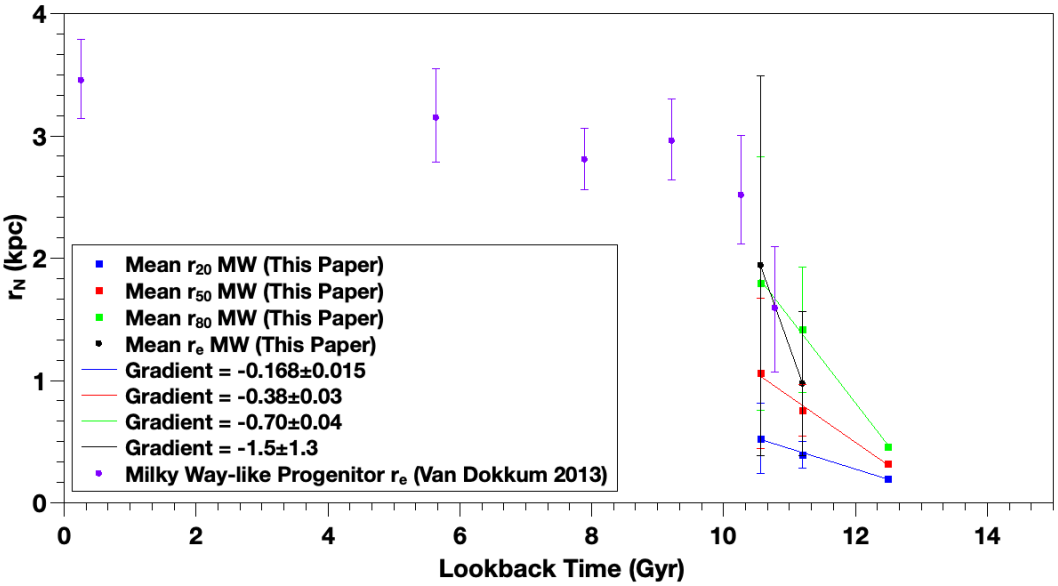

Figure 16. The sizes of our Milky Way-like progenitors as a function of lookback time for different radii (r_N) and their standard deviations, where blue is r_{20}, red is r_{50}, green is r_{80}, black is r_e from this paper and purple are the Milky Way-like progenitor r_e values from van Dokkum et al. (2013).

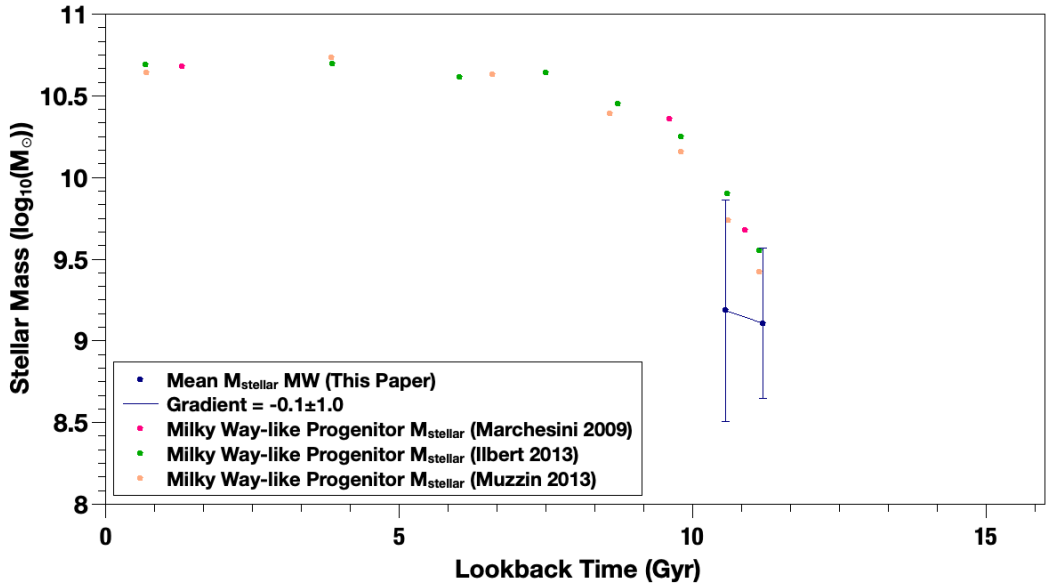

Figure 17. Mean stellar mass and standard deviations per redshift slice against lookback time (Gyr) for Milky Way-like progenitors, where navy is $M_{stellar}$ from this paper, and magenta, green and peach are the stellar masses of Milky Way-like Progenitors from Marchesini et al. (2009), Ilbert et al. (2013) and Muzzin et al. (2013) respectively (found in van Dokkum et al. (2013)).

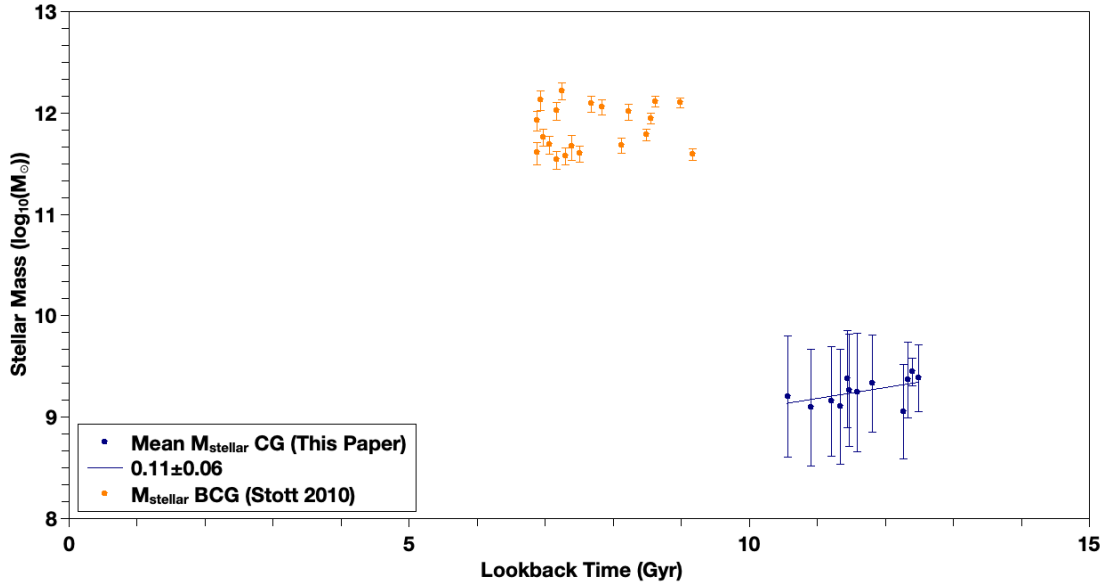

Figure 18. Mean stellar mass and standard deviations per redshift slice against lookback time (Gyr) for cluster-like progenitors, where navy is $M_{stellar}$ from this paper, and orange are the stellar masses of brightest cluster galaxies (BCGs) from Stott et al. (2010).

must experience a decrease in their average SFRs over time in order to become like the Milky Way. Overall, galaxies in our sample must have a stellar mass $M \sim 10^{10} M_{\odot}$ and SFR less than $10^{0.73}$ $M_{\odot} yr^{-1}$ in order to be classified as Milky Way progenitors.

6 DISCUSSION

In this section the results given in Section 4 are discussed, with the morphologies, SFR, AGN and Progenitor results discussed in Section 6.1, Section 6.2, Section 6.3 and Section 6.4, respectively.

6.1 Morphologies

Our values for the compactness ($C = 2.6 \pm 0.4$) and Sérsic index ($n = 1.9 \pm 2.2$) agree with those of Paulino-Afonso et al. (2018) for the same data set suggesting the majority of the galaxies in the SC4K data set are compact disks. Our VC mean results have a correlation of 0.64 with those of Paulino-Afonso et al. (2018) and this is likely due to a difference

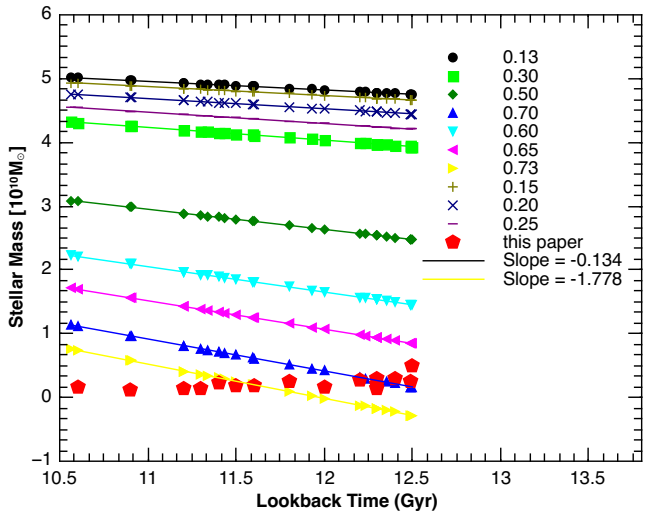

Figure 19. Stellar Mass against lookback time for galaxies that would evolve to have a mass equal to that of the Milky Way ($M = 6.43\times10^{10}$ $M_\odot$), meaning all lines would converge at LBT = 0. The values in the legend correspond to $\log_{10}(\frac{SFR}{M_\odot yr^{-1}})$. The average stellar mass (per bin) of galaxies in our sample are shown as red hexagons.

in interpretations of classification boundaries and does not stop conclusions being drawn from these results. The general growth of radius detailed in Section 4.1.2 agrees with the literature on evolution of LAEs, as does the roughly constant stellar mass, once selection bias has been taken into account. Our results show that at greater distances a much smaller number of less massive, low luminosity galaxies can be seen. This skews results upward at high redshifts and downwards at low redshifts for stellar masses and creating the slight arch seen in the VC mean and n results in Section 4.1.1 and Section 4.1.2. In the case of the whole of the SC4K data set, it can be suggested that as these galaxies are evolving they do not gain much stellar mass (Figure 7) to increase their radii, and rather flatten out from compact galaxies to less compact disky galaxies (Figure 4) as their 50% to 80% light radii regions are growing 1.75 times faster than their 20% to 50% light radii regions (see Section 4.1.2).

6.2 SFR

As outlined in Section 4.3 we found that SFR increases with increasing stellar mass and increasing redshift. This may be because galaxies with large stellar masses are able to draw in large quantities of gas, which enable them to form more stars than galaxies with lower stellar masses. Feedback may also be significant at lower redshifts, where galaxies have used up their original material, and in order to form new stars need to reuse material from stars which have gone supernovae. Only the more massive galaxies can do this, because galaxies with lower stellar masses have lower escape velocities, and therefore lose more stellar material. It is also possible that the most massive galaxies were able to become so massive due to having a large SFR.

At higher redshifts, the flux from galaxies becomes less and less, and hence we are only able to detect the most luminous galaxies (with highest SFR), and this therefore creates the selection bias mentioned previously. This is best shown in Figure 14, where we see that all SFRs increase with redshift, due to the flux limit. This particularly affects the upper limit results from radio and FIR stacking, which were obtained through the use of non-detections. Due to the stacking process, however low the flux in each case may have been, all that is obtained is the flux upper limit (relating to noise, rather than measurements). For increasing luminosity distance d_L (i.e. increasing redshift), a given flux will correspond to increasing luminosity L as $L \propto d_L^2$ (demonstrated by Equation 9 and Equation 13). SFR is in turn proportional to luminosity and this therefore leads to increasing SFR with increasing redshift.

In Figure 14, the results from Smit et al. (2012), Sobral et al. (2012a), and Reddy et al. (2008) are for the SFR function, not (like our results) for the average observational SFR of their sample. They are therefore typical SFRs for galaxies (SFR*). At lower redshift, we are able to observe a larger number of galaxies forming stars at a much lower rate than SFR*, and therefore our results are lower at this redshift (SFR < SFR*). However, in Figure 14 we find that at redshifts $z \sim 4-5$, the flux limit is such that we observe the typical SFR. Above this redshift, our selection bias pushes the observed SFRs up, and hence why our results continue to increase with redshift, whereas SFR* decreases.

In Figure 13 we see that the SFR predicted by Dutton et al. (2010) starts much lower than our results, and increases much more rapidly over the redshift range $z \sim 2.5-3.1$. Our results are generally higher than these model predictions due to the selection bias explained previously. We see that our data matches Dutton's predictions around stellar mass $M \sim 10^{10}$ $M_\odot$, after which Dutton et al. (2010) predictions continue to increase, whereas our results are roughly constant. This is due to the 'dust wall' that builds up in galaxies with higher stellar masses and higher SFRs. Between the redshift range $z \sim 4-5.5$ we compare to Salmon et al. (2015). These are median dust corrected UV SFR results from CANDELS (Cosmic Assembly Near-infrared Deep Extra-galactic Legacy Survey). In a similar fashion to Dutton et al. (2010), these results start much lower than our values, but then increase rapidly with stellar mass and become greater than our results. This shows the difference between dust-corrected and uncorrected SFRs.

6.3 AGN

Active galactic nuclei (AGN) are the highly active supermassive black holes at the centre of active galaxies, which emit strongly in the X-ray and radio wavelengths. The X-ray emissions of an AGN are produced by the accretion disc and the high energy particle jets of the supermassive black hole, with the radio emissions of the AGN also being produced by the jets as well as lobes of high energy particles fed by these jets. Previous studies have found that the activity of the AGN is also traced by the Lyα emissions of the dust surrounding the black hole, with the brightest LAEs typically being AGN (e.g. Sobral et al. 2018a,b). Thus, as Lyα traces both AGN activity and SFR, the emissions from the AGN will contaminate the emissions due to star formation

within the active galaxy, the SFR of AGN cannot be calculated using the Lyα emission lines. Additionally, as X-ray and radio emissions also trace the AGN activity and particle jets, the SFR cannot be calculated from those emission lines. The high Lyα emissions of AGN also cause them to appear bluer and more luminous than star forming LAEs, which causes the SED fitting of an AGN light curve to give a smaller stellar mass value for the AGN due to the AGN emissions causing the galaxy to appear to contain more O and B type stars. Despite being unable to investigate the mass and SFR trends of the AGN in the SC4K catalogue, we are able to investigate the trends in the fraction of AGN in the SC4K catalogue and the BHAR of these AGN with redshift, Lyα luminosity and their morphologies.

The AGN in our sample are found to be highly luminous with the fraction of X-ray and radio sources in our catalogue increasing with the log of the Lyα luminosity, which is consistent with results found in Matthee et al. (2017) and the trend found with spectroscopic data in Sobral et al. (2018a) and Wold et al. (2014). Thus we conclude that AGN are typically the brightest LAEs across all redshifts.

The fraction of AGN in our catalogue at different redshifts shows that the abundance of AGN in the universe peaks at a redshift $z \sim 3$, just before the peak in star formation rate at $z \sim 2$, and decreases with increasing redshift. This suggests that there may be a relation between an increase in the activity of the supermassive black hole of a galaxy and the burst of star formation within the galaxy, as in Madau & Dickinson (2014), producing the SFRD peak at cosmic noon.

From the morphology data, obtained for each galaxy in the catalogue, it was found that the AGN identified are point-like compact galaxies, with most of the light emitted by the active galaxy produced within ~ 4.5 kpc of the geometric centre of the galaxy. The light radii of these AGN were found to decrease with increasing BHAR (see Figure 11), indicating the majority of the emissions of an active galaxy being produced by the accretion disk of the supermassive black hole itself. Thus AGN appear as bright, compact, point-like sources for which little of the surrounding galaxy can be observed due to the high luminosity of the supermassive black hole.

At high redshifts only the brightest galaxies can be observed. Our sample size at high redshifts is small, a high fraction of which are AGN likely due to the high luminosity of AGN allowing them to be observed at these high redshifts. This may have produced a sample bias in the fraction of observed galaxies at high redshifts, and the peak in the fraction of AGN at high redshifts may be an artefact of this sample bias. Thus deeper searches at high redshifts are required to further investigate the trends presented in this paper.

6.4 Progenitors

The radii results from our Milky Way-like progenitor data set fit well to results from van Dokkum et al. (2013) and thus it can be expected that they will follow trends in this data and decrease their speed of growth with time. The stellar mass results also fit, within error bars, to the data from Marchesini et al. (2009), Ilbert et al. (2013) and Muzzin et al. (2013) and thus we can also expect them to follow the trend. While our Milky Way-like progenitor Sérsic index values may technically agree with the data within errors bars, said error bars are far too large for these results to offer any meaningful conclusions.

As the selection bias present in the SC4K catalogue works in our favour with cluster-like progenitors, since they are the most luminous of our three categories, we can identify more potential cluster-like progenitors and increase the range of lookback times we can study. As with all of our results the radii of the cluster-like progenitors are increasing in general, however in their case they increase sharply just before a lookback time of ~ 10 Gyrs which is the point at which De Lucia & Blaizot (2007) predicts that these galaxies will start growing via mergers. Practically no compact or disky classifications are found at lookback times approaching ~ 10 Gyrs and a large number of point-like and irregular (which includes merger) classifications are. This implies, along with the spike in radii size at a similar time, that the cluster-like progenitors in the SC4K are beginning to merge with other galaxies and grow through that mechanism rather than stellar mass production. Similarly to its effect as described in Section 4.1, the selection bias present in the SC4K catalogue is likely to be the cause of the downward trend seen in Figure 18. This being said BCGs are formed mostly by galaxy mergers after lookback time ~ 10 Gyrs and thus it might not necessarily be expected to see a trend that could be extrapolated to match to the Stott et al. (2010) data as they have likely merged many times since lookback time ~ 10 Gyrs. In contrast to the the Milky Way-like progenitors, cluster-like progenitor Sérsic index values show an increase that agrees with results from Nelson et al. (2002). While the errors are large the spread of data points is mirrored in the data from Nelson et al. (2002) suggesting that, as bright cluster galaxies are formed through mergers, a large range of n values is to be expected but that on average cluster-like progenitor Sérsic indexes increase with age.

In Section 5.2.2 we found that the majority of galaxies in the SC4K sample must decrease their SFRs over cosmic time in order to have a stellar mass similar to that of the Milky Way at present. The results of this paper suggest that SFR decreases with decreasing redshift in the range $z \sim 6-2$ (see Figure 14), which suggests that a significant portion of our sample has the potential to have a mass similar to that of the Milky Way today. Therefore galaxies at redshifts $\sim 6-2$ with stellar mass M $\sim 10^{10}$ M$_\odot$ and SFR $\sim 10^{0.73}$ M$_\odot$yr^{-1} (see Section 5.2.2) will be Milky Way progenitors in terms of stellar mass.

However, the results from Smit et al. (2012), Sobral et al. (2012a), and Reddy et al. (2008) (see Figure 14) show that the typical SFR (SFR*) increases with decreasing redshift from $z \sim 6-2$. Therefore, according to these results, there must be a steep decline in SFR between $z \sim 2-0$ in order for galaxies in the SC4K sample to be Milky Way progenitors in terms of stellar mass.

7 CONCLUSIONS

In this paper we analyse a sample of ~ 4000 high redshift Lyα emitters (LAEs), specifically looking at their morphologies, SFRs, and AGN activity. Finally we also attempt to

predict the properties of local galaxies at an earlier time. Our main results are:

- The SC4K catalogue is mostly made up of compact disky galaxies (average $n = 1.9 \pm 2.2$) taking up a $69 \pm 4\%$ share of the overall total with the complete set having a mean compactness, $C = 2.6 \pm 0.4$.
- The average $SFR_{Ly\alpha}$ across our whole sample (excluding AGN) is $16.5 \pm 0.3\ M_{\odot}yr^{-1}$, and an average SFR_{UV} (uncorrected for dust) of $4.55 \pm 0.08\ M_{\odot}yr^{-1}$. In general, our higher redshift LAEs present higher SFRs, a consequence of a higher Lyα luminosity limit. Our radio and FIR stacking measurements form upper limits to the SFR in our sample, as they are calculated from non-detections. They are mostly found to be between $100 - 1000\ M_{\odot}yr^{-1}$.
- SFR is also found to increase with increasing stellar mass. In addition, we also find a characteristic 'bump' in SFR at $M \sim 10^{9.3}\ M_{\odot}$ at $z \sim 2.5$. This bump is shown to progressively move towards higher stellar masses at higher redshifts. We find that the dust content varies with stellar mass in galaxies, and we find that this peaks at stellar masses $M \sim 10^{10}\ M_{\odot}$, at which point a 'dust wall' is likely preventing a large fraction of Lyα and UV photons from escaping (Sobral et al. 2018b).
- In the SC4K catalogue 303 AGN were identified, using X-ray and radio emissions, with BHARs from $\sim 0.03\ M_{\odot}\ yr^{-1}$ to $\sim 3.3\ M_{\odot}\ yr^{-1}$.
- We found that the percentage fraction of AGN in the SC4K catalogue increases with increasing Lyα luminosity and decreases with increasing redshift, peaking at a redshift of $z \sim 3$ just before the peak in SFRD of cosmic noon.
- The AGN morphologies found show that AGN are typically point-like elliptical galaxies, with the Sérsic and percentage light radii decreasing with increasing BHAR.
- SC4K LAEs start off as compact galaxies and then flatten out into less compact disks, without a large amount of stellar mass growth, their 50% to 80% light radii range growing ~ 1.75 faster than their 20% to 50% light radii range.
- We show that LAEs in the range $z \sim 2 - 6$ of different luminosities may evolve into galaxies with a range of masses in the local Universe.
- LAEs that will likely evolve into Milky Way-type galaxies in the local universe are rare in the SC4K data set however they do seem to follow established galaxy evolutionary trends.
- The majority of high redshift LAEs are likely to become the massive elliptical galaxies at the cores of the brightest cluster galaxies (BCGs). Our analysis shows that these LAEs show signs of beginning their mass and radii evolution via merging with other galaxies at a lookback time of ~ 10 Gyrs with little stellar mass evolution before this point.
- In terms of stellar mass, we found that galaxies in the redshift range $z \sim 2 - 6$ need a stellar mass $M \sim 10^{10}\ M_{\odot}$ and a SFR $\sim 5.4\ M_{\odot}yr^{-1}$ in order to be classified as a Milky Way progenitor. However, given that the sample average SFR $\approx 12.6\ M_{\odot}yr^{-1}$, galaxies in the early Universe must experience a decrease in their average SFRs over time in order to become like the Milky Way.

The high redshift LAEs in the SC4K data set provide an insight into the origins of some of the most massive galaxies in the local Universe as well as a look at how some early Milky Way-like progenitors may have looked. Further light could be shed on this latter group of early galaxies if higher resolution and deeper field data could be provided for the range $z \sim 2 - 6$ as this would allow a greater number of less massive galaxies to be studied which would result in more Milky Way-like (and possibly some dwarf-like) progenitor sources to study. This extra data would also shed more light on the further evolution of cluster-like progenitors (especially if the survey could be extended to $z \sim 1.5$) as well as reducing the effect of the selection bias that has been a feature in many of our results. Ultimately it would be ideal for a deep field study to be carried out on a section of sky and a subsequent catalogue, similar to SC4K, to be created over a full range of wavelengths and from $z \sim 0 - 8$ or further, in order to get a clearer picture of the evolutionary path of all types of galaxy found in the local Universe.

ACKNOWLEDGMENTS

The authors are thankful to the COSMOS survey teams, PHYS369 and the PIGEONS survey. We have benefited immensely from the public available programming language Python, including NumPy & SciPy (Van Der Walt et al. 2011; Jones et al. 2001), Matplotlib (Hunter 2007), Astropy (Astropy Collaboration et al. 2013), the Topcat analysis programs (Taylor 2013) and QtiPlot (Vasilef I 2009) and the fitting program Galfit (Peng et al. (2002); Peng et al. (2010)). This research has made use of the COSMOS Cutouts Service provided by the NASA/IPAC IR Science Archive.

REFERENCES

Astropy Collaboration et al., 2013, A&A, 558, A33
Bacon et al., 2015, A&A, 575, A75
Calhau J., 2019, MNRAS, submitted
Caon N., Capaccioli M., D'Onofrio M., 1993, MNRAS, 265, 1013
Casey C. M., 2012, MNRAS, 425, 3094
Cassata P., et al., 2010, ApJ, 714, L79
Charlot S., Fall S. M., 1993, ApJ, 415, 580
Civano F., et al., 2016, ApJ, 819, 62
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
Delhaize J., et al., 2017, A&A, 602, A4
Dutton A. A., van den Bosch F. C., Dekel A., 2010, MNRAS, 405, 1690
Elvis M., et al., 2009, ApJS, 184, 158
Gruppioni C., et al., 2013, MNRAS, 432, 23
Haardt F., Maraschi L., 1991, ApJ, 380, L51
Hagen A., et al., 2016, arXiv e-prints, p. arXiv:1610.01163
Hashimoto et al., 2017, A&A, 608, A10
Hathi N. P., Le Fèvre O., 2016, in Kaviraj S., ed., IAU Symposium Vol. 319, Galaxies at High Redshift and Their Evolution Over Cosmic Time. pp 22–25 (arXiv:1603.05992), doi:10.1017/S174392131501039X
Hayes M., Schaerer D., Östlin G., Mas-Hesse J. M., Atek H., Kunth D., 2011, ApJ, 730, 8
Hogg D. W., 1999, arXiv Astrophysics e-prints,
Hunter J. D., 2007, Computing In Science & Engineering, 9, 90
Ilbert O., et al., 2009, ApJ, 690, 1236
Ilbert O., et al., 2013, A&A, 556, A55
Jones E., Oliphant T., Peterson P., et al., 2001, SciPy: Open source scientific tools for Python, http://www.scipy.org/
Karim A., et al., 2011, ApJ, 730, 61
Kennicutt Jr. R. C., 1998, ARA&A, 36, 189

Khostovan A. A., Sobral D., Mobasher B., Best P. N., Smail I., Stott J. P., Hemmati S., Nayyeri H., 2015, MNRAS, 452, 3948
Khostovan A. A., et al., 2018, arXiv:1811.00556,
Koekemoer A. M., et al., 2007, ApJS, 172, 196
Laursen P., Duval F., Östlin G., 2013, ApJ, 766, 124
Lehmer B. D., et al., 2013a, ApJ, 765, 87
Lehmer B. D., et al., 2013b, ApJ, 765, 87
Lilly S. J., Fèvre O. L., Hammer F., Crampton D., 1996, ApJ, 460
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Malhotra S., Rhoads J. E., Finkelstein S. L., Hathi N., Nilsson K., McLinden E., Pirzkal N., 2012, ApJ, 750, L36
Marchesini D., van Dokkum P. G., Förster Schreiber N. M., Franx M., Labbé I., Wuyts S., 2009, ApJ, 701, 1765
Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, MNRAS, 351, 169
Matthee J., Sobral D., Best P., Smail I., Bian F., Darvish B., Röttgering H., Fan X., 2017, MNRAS, 471, 629
McMillan P. J., 2011, MNRAS, 414, 2446
Muzzin A., et al., 2013, ApJ, 777, 18
Nelson A. E., Simard L., Zaritsky D., Dalcanton J. J., Gonzalez A. H., 2002, ApJ, 567, 144
Netzer H., 1989, Comments on Astrophysics, 14, 137
Neufeld D. A., 1991, ApJ, 370, L85
Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
Oteo I., Sobral D., Ivison R. J., Smail I., Best P. N., Cepa J., Pérez-García A. M., 2015, MNRAS, 452, 2018
Paulino-Afonso A., et al., 2018, MNRAS, 476, 5479
Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266
Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2010, AJ, 139, 2097
Pirzkal N., Malhotra S., Rhoads J. E., Xu C., 2007, ApJ, 667, 49
Reddy N. A., Steidel C. C., Pettini M., Adelberger K. L., Shapley A. E., Erb D. K., Dickinson M., 2008, ApJS, 175, 48
Rowan-Robinson M., et al., 2016, MNRAS, 461, 1100
Salmon B., et al., 2015, ApJ, 799, 183
Salpeter E. E., 1955, ApJ, 121, 161
Santos S., Sobral D., Matthee J., 2016, MNRAS, 463, 1678
Sérsic J. L., 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41
Shapley A. E., Steidel C. C., Erb D. K., Reddy N. A., Adelberger K. L., Pettini M., Barmby P., Huang J., 2005, MNRAS, 626, 698
Shibuya T., Ouchi M., Harikane Y., Nakajima K., 2019, ApJ, 871, 164
Shields G. A., 1999, PASP, 111, 661
Smit R., Bouwens R. J., Franx M., Illingworth G. D., Labbé I., Oesch P. A., van Dokkum P. G., 2012, ApJ, 756, 14
Smolčić V., et al., 2017, A&A, 602, A6
Sobral D., Matthee J., 2019, A&A, 623, A157
Sobral D., Best P. N., Matsuda Y., Smail I., Geach J. E., Cirasuolo M., 2012a, MNRAS, 420, 1926
Sobral D., Smail I., Stott J. P., Best P. N., Geach J. E., Cirasuolo M., Matsuda Y., Kurk J., 2012b, MNRAS, 428, 1128
Sobral D., Matthee J., Darvish B., Schaerer D., Mobasher B., Röttgering H. J. A., Santos S., Hemmati S., 2015, ApJ, 808, 139
Sobral D., Santos S., Matthee J., Paulino-Afonso A., Ribeiro B., Calhau J., Khostovan A. A., 2018a, MNRAS, 476, 4725
Sobral D., et al., 2018b, MNRAS, 477, 2817
Somerville R. S., Dave R., 2015, ARA&A, 53, 51
Stark D. P., 2016, ARA&A, 54, 761
Stott J. P., et al., 2010, ApJ, 718, 23
Taylor M., 2013, Starlink User Note, 253
Van Der Walt S., Colbert S. C., Varoquaux G., 2011, Computing in Science & Engineering, 13, 22
Vika M., Bamford S. P., Häußler B., Rojas A. L., Borch A., Nichol R. C., 2013, MNRAS, 435, 623
Wisotzki et al., 2016, A&A, 587, A98
Wold I. G. B., Barger A. J., Cowie L. L., 2014, ApJ, 783, 119
van Dokkum P. G., et al., 2013, ApJ, 771, L35

This paper has been typeset from a TEX/LATEX file prepared by the author.